The Dissertation Committee for Brent Laurence Hughes Certifies that this is the approved version of the following dissertation:

Using the Neural Level of Analysis to Understand the Computational Underpinnings of Positivity Biases in Self-Evaluation

Committee:

Jennifer S. Beer, Supervisor

Samuel D. Gosling

Lisa A. Neff

Alison A. Preston

William B. Swann, Jr.

Using the Neural Level of Analysis to Understand the Computational Underpinnings of Positivity Biases in Self-Evaluation

by

Brent Laurence Hughes, B.S.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

May 2012

Dedication

This dissertation is dedicated to my parents, Maria Elena and Laurence Hughes, to my sisters, Natalie and Jessica, and to my closest friends, for believing in me and providing unconditional support and encouragement, and to Richard Mann for showing me the way.

Acknowledgements

There is no way to fully express the gratitude I feel for the people who have been a part of my life and provided endless support and encouragement during my time in Austin. Reflecting on my time here, I feel fortunate and blessed to have a group of amazing people as mentors, colleagues, and friends.

First, thank you to my advisor, Jennifer Beer, for her support and guidance throughout graduate school. From the very beginning, Jenni has been a fiercely dedicated advisor, providing endless amounts of time and challenges to help me grow and develop as a thinker and as an interdisciplinary scientist. She pushed me to always think critically about any task at hand and think about problems in new ways. Her willingness to take risks and think creatively led me to discover new areas of inquiry within the domains of self-knowledge and social cognition.

Second, thank you to my dissertation committee members, Alison Preston, Bill Swann, Sam Gosling, and Lisa Neff. Thank you all for your insightful comments and feedback on my research, for challenging me to make novel connections with other literatures, and for volunteering so much of your time to meet with me. Ali, thank you for our numerous meetings to discuss a wide array of topics ranging from my own research to the state of the field, and for all of your valuable career advice.

Third, thank you to the entire Social Personality area. I've been fortunate to be in the company of faculty and students who are incredibly diverse in their research interests, who share a mutual respect for each other's work, and who are always open to have animated conversations about research ideas and life. Being in such a rich and stimulating environment has influenced me as a researcher and provides a benchmark for the kind of environment I hope to find in the future. Thanks to Jamie Pennebaker for always asking,

"Why should we care, what does this all mean?" and to Bob Josephs for providing a unique combination of humor and insight to any interaction.

I'd like to extend special thanks to the past and present members of the lab, including Jamil Bhanji, Pranj Mehta, Hani Freeman, Gili Freedman, Taru Flagan, Janell Fetterolf, and David Chester. Immense gratitude goes to Jamil for being an incredible colleague and wonderful friend. Thank you for dealing with my many questions and doubts, and for providing encouragement and support through the many challenges of graduate school. I will miss having you as a lab-mate, our JPs runs, our fantasy soccer teams, and our conversations. Life in graduate school would not have been the same without you, and I am grateful that you, Emily and Gabriel made me feel like a part of your family. Pranj and Hani, thank you for paving the way and providing a big picture perspective during graduate school and beyond. Gili, Taru, Janell, and Dave, you are all like younger academic siblings to me, thank you for completing my academic family and being obliged to laugh at all of my jokes.

A heartfelt thank you to the amazing friends I've made by being a part of this exciting field. Thank you to Michael Buhrmester for being my office mate and conference roommate these 5 years of grad school. You are a gentleman and a scholar, a gem of a human being, and I hope our paths continue to cross in the future. Thank you to Simine Vazire for being a research role model and friend during these years of grad school. I am grateful for your invitation to give a talk at WashU and our time spent in Princeton eating chocolate croissants and talking about the self. Thank you to the crew at Columbia and New York, in particular Kevin Ochsner, Tor Wager, and Ed Smith for providing amazing mentorship before my graduate school days, and to Sam Gershman, Ivan Barenboim, Jamil Zaki, Hedy Kober, Matthew Davidson, Josh Davis, Lauren Kaplan, Rachel Insler, Lauren Atlas, Kirstin Appelt, Diego Berman, and the rest of the

New York gang that have continued to be in my life during my own adventures in graduate school. A huge thank you to the many dear friends I've made at conferences, workshops, and the academic life (you know who you are).

A warm and heartfelt thank you to all of my Austin friends that made this place a wonderful home that I will greatly miss. Special thanks to A. Ross Otto, J. Grant Loomis, Emily Brownell, Lisa Gulessarian, Shannon Nagy, Nick Gaylord and Micah Goldwater, for our many camping trips, food explorations, racquetball tournaments, bike rides, and other adventures it would be imprudent of me to reminisce about here. Special thanks to Bridget Mouton, Amanda Mullee, Lee Kirby Webster, and Boone, for many wonderful evenings filled with cooking and wine, picnics, swimming trips, and conversations. Yet another thanks to Jamil, Emily, and Gabriel for making me feel a part of your family, for our Thanksgiving traditions, and for your endless support. Lastly, thanks to Brian Sullivan, Cindy Chung, Scott Liening, Molly Ireland, Matt Brooks, Tyler Davis, Jose Barragan and FC Subaltern, and the many people that I've been lucky to call friends here.

Thank you to Sally, for her endless support through this challenging journey. She shared the joys and pains of graduate school day after day, from being overjoyed at my initial admission to grad schools and having my first paper accepted, to hearing me agonize over failed experiments and rejected papers. She has reminded me to eat, sleep, bathe, and take breaks during my darkest hours, and provided much-needed happiness, stability, and support to my life. I can't imagine what this journey would have been like without you. Finally, I would like to thank my entire family for their support and encouragement. To my parents, Maria Elena and Laurence, to my sisters Natalie and Jessie, to my nieces Sophia and Isabella, to my aunt Alicia, to my brothers from other mothers Amjad Majid and Zacharias Greenberg, and to my mentor and role model, Richard Mann. I am forever grateful.

Using the Neural Level of Analysis to Understand the Computational Underpinnings of Positivity Biases in Self-Evaluation

Brent Laurence Hughes, Ph.D.

The University of Texas at Austin, 2012

Supervisor: Jennifer S. Beer

Decades of research have demonstrated that people sometimes provide self-evaluations that emphasize their most flattering qualities. Different theoretical accounts have been offered to explain the mechanisms underlying positively-biased self-evaluation. Some researchers theorize that positively-biased self-evaluations arise from a self-protection motivation because positivity biases increase in situations of heightened self-esteem threat. Alternative views question whether self-protection motivation is a necessary or even dominant source of positivity bias by demonstrating that positively-biased self-evaluations occur even when threat is not heightened, and that a general judgment approach leads to positivity biases in some domains but also to negativity biases in other domains. One reason for this gap in knowledge is that behavioral measures are limited in their ability to resolve whether the processes underlying positively-biased self-evaluation are the same or different depending on contextual motivators. Neuroimaging methods are well suited to examine whether different mechanisms underlie similar behaviors, specifically similar positively-biased responses in different contexts. The four studies presented here explore the neural mechanisms of

positively-biased self-evaluation by first identifying a core set of neural regions associated with positivity bias (Study 1A and 1sB), examining whether a heightened self-protection motivation changes the engagement of those neural systems (Study 2), and specifying the precise mechanisms supported by those regions (Study 3). Studies 1A and 1B revealed evidence for a neural system comprised of medial and lateral orbitofrontal cortex (OFC) and, to a lesser extent dorsal anterior cingulate (dACC) that was modulated by positivity bias. Study 2 found that a heightened self-protection motivation changes the engagement of medial OFC in positively-biased self-evaluation. Finally, Study 3 found evidence that medial OFC may support a common mechanism in positively-biased judgment that is implemented differently as a function of the motivational context. Taken together, these studies represent a first step toward developing a neural model of positively-biased self-evaluation. The findings provide some preliminary evidence that positivity biases may represent distinct processes in different motivational contexts. This dissertation sets the stage for future work to examine how specific positively-biased cognitive mechanisms may be supported by specific neural systems and computations as a function of motivational contexts.

Table of Contents

List of Tables

List of Figures

INTRODUCTION

Despite the admonition from the Oracle at Delphi to "know thyself," people's evaluations of themselves are sometimes flawed and in a remarkably systematic manner. Specifically, decades of research suggest that people sometimes evaluate themselves in a more flattering manner than what external criteria would suggest across a wide variety of domains, including their personality characteristics, the knowledge they possess, and even their actual behavior (e.g., Alicke, 1985; Dunning, Meyerowitz, & Holzberg, 1989; Gosling, John, Craik, & Robins, 1989; Klayman et al., 1999; Paulhus et al., 2003; Sedikides & Gregg, 2008; Taylor & Brown, 1988). One point of debate is whether the computational underpinnings of these positivity biases in self-evaluation are the same or different as a function of different contextual motivators. The gap in knowledge is partially due to the limitations of behavioral measures to resolve whether similar or different underlying processes support positively-biased self-evaluations in different motivational contexts. The neural level of analysis can provide information about the mechanisms that underlie similar positively-biased responses in different motivational contexts. This dissertation represents a step away from underspecified behavioral comparisons of positively-biased self-evaluations in different contexts by examining the neural mechanisms that underlie positively-biased self-evaluations in different motivational contexts.

Empirical Evidence of Positivity Biases in Self-Evaluation

One of the most robust examples of positivity biases in self-evaluation is the "better than average" effect, that is, the tendency for the majority of people evaluate themselves as having more desirable characteristics and fewer undesirable characteristics than their average peer (Alicke, 1985; Brown, 1986; Chambers & Windschitl, 2004; Dunning et al., 1989; Moore & Small, 2007; Weinstein, 1980). People also extend these better-than-average evaluations to their close others (e.g., romantic partners, best friends), but not to their non-close others (e.g., acquaintances, peers)(Brown, 1986; Buunk & Van Yperen, 1991; Gagne & Lydon, 2001; Murray & Holmes, 1997; Suls et al., 2002; Van Lange & Rusbult, 1995). Although each person may have some unique characteristics, an average peer is also likely to have some unique characteristics. Therefore, it is logically improbable that the majority of people in a randomly selected sample would be better than their average peer across a large number of traits (Chambers & Windschitl, 2004; Taylor & Brown, 1988). Instead, it would be expected that evaluating the self across a large number of traits should be centrally distributed around the average peer (Chambers and Windschitl 2004; Taylor & Brown, 1988). Therefore, the tendency for people to evaluate themselves as better than their average peer across a large number of traits provides compelling evidence that people may have positively-biased views of themselves relative to others. These better than average effects are not constrained to samples of college students, and predict real-world outcomes across a variety of domains (e.g., health, education, business, law, and even college professors: Babcock &

Loewenstein, 1977; Cooper, Woo, & Dunkelberg, 1988; Cross, 1977; Dunning, Heath, & Suls, 2004; Larwood, 1978; Loftus & Wagenaar, 1988; Odean, 1998; Rutter, Quine, & Albery, 1998).

In addition to providing positively-biased evaluations of their personality characteristics, people also tend to claim more knowledge about scholarly concepts than they really have in order to appear intelligent (overclaiming bias: Paulhus et al., 2003; Paulhus & Harms, 2004; Phillips & Clancy, 1972; Randall & Fernandes, 1991; Stanovich & Cunningham, 1992). Research on overclaiming has shown that people tend to inflate how much they claim to know about real scholarly concepts, and sometimes even claim to know nonexistent information (Paulhus et al., 2003). Research using this task has applied signal detection theory (SDT; Green & Swets, 1966; Macmillan & Creelman, 1991) techniques to model people's tendency to make exaggerated claims of knowledge (i.e., decision threshold (*c*)). From a SDT perspective, decision thresholds provide a measure of overclaiming because they are theorized to reflect how strong a sense of familiarity is needed in order to claim knowledge. For example, people who are more positively-biased tend to claim as much knowledge as possible and may accomplish that goal by considering a very weak sense of familiarity as indicative of actual knowledge. The people who are most likely to consider a very weak sense of familiarity as indicative of actual knowledge when claiming to know scholarly concepts are narcissists and those motivated to deceive others into seeing them in a positive light (Bing, Kluemper, Davison, Taylor & Novicevic, 2011; Paulhus et al., 2003; Randall & Fernandes, 1991; Tracy, Cheng, Robins & Trzesniewski, 2009).

Lastly, people also tend to provide positively-biased evaluations of their performance and actual behavior. For example, people often report levels of confidence in their performance on a variety of tasks (e.g., answering trivia questions, predictions about the future, medical and clinical diagnoses) that exceeds their actual task performance (overconfidence bias, e.g., Critcher & Dunning, 2009; Fischhoff et al., 1977; Klayman et al., 1999; Koriat et al., 1980; Oksam et al., 2000; Oskamp, 1965). In addition, people sometimes provide self-evaluations of their actual behavior that are more favorable than the evaluations of outside observers (e.g., Colvin, Block, & Funder, 1995; Gosling et al., 1998; John & Robins, 1994; Paulhus, 1998; Robins & Beer, 2001). For example, people remember performing more desirable behaviors during a group discussion task than objective observers can later identify (Gosling et al., 1998). Moreover, people tend to attribute positive outcomes to the self and dismiss negative outcomes to factors outside the self (e.g., Bradley, 1978; Campbell & Sedikides 1999; Jones & Nisbett, 1971; Miller & M. Ross, 1975; L. Ross, 1977; Zuckerman, 1979). Taken together, research suggests that people tend to provide positively-biased evaluations when comparing themselves to their peers, evaluating their scholarly knowledge, and evaluating their actual behavior.

While decades of research have provided evidence that people tend to have positively-biased views of themselves and certain other people across a variety of domains, it has not answered one essential question: do positivity biases represent a single phenomenon or do positivity biases arise from distinct processes in different motivational contexts? One point of debate is whether positively-biased judgment arise

from a motivation to protect the self, a general judgment strategy that is not tethered to self-protection concerns, or both depending on the motivational context (Alicke & Govorun, 2005; Chambers & Windschitl, 2004; Dunning et al., 1989; Kunda, 1990; Moore & Small, 2007; Sedikides & Gregg, 2008; Taylor & Brown, 1988).

Different Explanatory Approaches for How Positivity Biases Are Accomplished

SELF-PROTECTION PERSPECTIVE

Some researchers have proposed that positively-biased self-evaluations are best explained by a motivation to protect the self (e.g., Alicke et al., 1995; Dunning, 1995; Kunda, 1990; Sedikides & Gregg, 2008; Swann, 2011; Taylor & Brown, 1988). From a self-enhancement perspective, most people are motivated to protect the self and promote feelings of self-worth, and partially accomplish that motivation by evaluating their personality characteristics, knowledge, and behaviors in self-serving ways (e.g., Taylor & Brown, 1988; Sedikides & Gregg, 2008). Another self-protection perspective suggests that people are motivated to protect the consistency of self-views (Swann, 2011). From a self-verification perspective, people partially accomplish that motivation by seeking self-verifying feedback and providing evaluations that are consistent with their firmly-held self-views (Kwang & Swann, 2010; Swann, 2011). Both self-protection perspectives agree that people with positive self-views are motivated to protect their positively-held self-views and therefore evaluate themselves in a positive manner (Kwang & Swann, 2010; vanDellen et al., 2011). The two perspectives differ in their predictions about

individuals with negative self-views. While the self-enhancement perspective predicts that individuals with negative self-views will strive for positivity much like individuals with positive self-views (Taylor & Brown, 1988; Sedikides & Gregg, 2008), the self-verification perspective predicts that individuals with negative self-views will strive to confirm their negative self-views (Kwang & Swann, 2010; Swann, 2011). As most people have positive self-views (Gray-Little, Williams, & Hancock, 1997; Koole et al., 2001; Twenge & Campbell, 2008) and both perspectives have similar predictions of individuals with positive self-views, discussion of the self-protection perspective will center on individuals with positive self-views. Support for the self-protection explanation comes from two lines of research.

First, researchers have suggested that positivity biases are elicited by a self-protection motivation because these evaluations often occur in situations involving heightened or explicit threats to the self (e.g., Alicke, 1985; Brown, 2012; Dunning, 1995; Sedikides & Gregg, 2008; vanDellen et al., 2011). For example, when people consider certain attributes to be more desirable, important, or related to success, they tend to evaluate themselves more favorably on those attributes (Brown, 2012; Dunning, 1995; Kunda & Sanitioso, 1990; Miller, 1976; Paulhus et al., 2003; Story, & Dunning, 1998). Most people may be motivated to view themselves as being competent, productive and capable of attaining positive life outcomes, and may partially accomplish that motivation by evaluating themselves favorably on success-related attributes. For example, better than average judgments are greater for traits that are considered to be important (Brown, 2012), and overclaiming is greater for people who are motivated to deceive others into

viewing them in a positive light and for narcissistic individuals who have a high need to make a good impression on others (Paulhus et al., 2003). Similarly, people tend to recall performing behaviors more frequently when the behaviors are described as being predictive of success and other desirable outcomes (Markus & Kunda, 1986; Sanitioso, Kunda, & Fong, 1990; Ross, McFarland, & Fletcher, 1981). People are also more enthusiastic about receiving feedback on important attributes, but only if they believe (or are led to believe) that they possess those important attributes (Dunning, 1995). Admitting to shortcomings on success-related attributes has the potential to threaten the self and undermine a self-protection goal.

In fact, explicit threats to the self have been shown to increase positively-biased self-evaluations, presumably as a way to compensate for the threat and protect self-worth (Brown, 2012; Dunning & Beauregard, 2000; Campbell & Sedikides, 1999; vanDellen et al., 2011). Threat refers to negative feedback about personality, academic competence, social skills, or interpersonal relationships that challenges favorable self-views (Baumeister, Heatherton, & Tice, 1993; Leary et al., 1998, 2009; vanDellen et al., 2011). Threats to the self tend to decrease self-esteem and therefore motivate people to compensate for the threat in a variety of ways (Campbell & Sedikides, 1999; Crocker & Park, 2004; Leary et al., 1998; vanDellen et al., 2011). For example, explicit threats increase the extent to which people view themselves as better than their average peer (Beer, Chester, & Hughes, forthcoming; Brown, 2012; Vohs & Heatherton, 2004), downplay their negative qualities and exaggerate their positive qualities (Baumeister & Jones, 1978; Brown & Smart, 1991; Greenberg & Pyszczynksi, 1985; Schneider, 1969),

take personal credit for successes while attributing failures to factors outside the self (Blaine & Crocker, 1993; Campbell & Sedikides, 1999; Shrauger & Lund, 1975), report more optimism and self-confidence about their ability to succeed in the future (McFarlin & Blascovitch, 1981; Josephs, Markus, & Tafarodi, 1992), and shift attention to positive characteristics and core values (Aronson, Blanton, & Cooper, 1995; Dodgson & Wood, 1998; Spencer, Josephs, & Steele, 1993; Steele, Spencer, & Lynch, 1993). Therefore, explicit threats to the self may engage a strong motivation to protect the self and bring about positively-biased self-evaluations to compensate for the threat.

Another reason researchers have suggested that positivity biases arise from a self-protection motivation is because preemptively bolstering the self by affirming core values and important aspects of the self reduces positively-biased self-evaluations. Research suggests that self-affirmation may temporarily satisfy a motivation to protect the self (Crocker & Park, 2004; Sedikides & Gregg, 2008; Steele, 1988; Tesser, 2000). Self-affirmation does not necessarily resolve the initial threat but rather bolsters the self by refocusing attention to valued aspects of the self. Preemptively bolstering the self through self-affirmation has been shown to reduce the tendency for people to exaggerate their performance (Gramzow & Willard, 2006), the tendency for people to reject critical feedback (Kumashiro & Sedikides, 2005; Sherman & Cohen, 2002; Sherman, Nelson, & Steele, 2000), and the tendency for people to boost their positivity by making downward social comparisons (e.g., comparing the self to incompetent others)(Fein & Spencer, 1997; Spencer, Fein, & Lomore, 2001). Taken together, research suggests that positivity biases in self-evaluations may arise from a motivation to protect the self because these

evaluations are increased in situations of heightened or explicit threat and reduced when self-affirmation preemptively bolsters the self from threat.

ALTERNATIVE PERSPECTIVE

While the bulk of research has typically offered a self-protective motivation to explain positivity biases in self-evaluations, an alternative perspective has recently suggested that there are reasons for looking beyond a strictly self-protective explanation to consider alternative or parallel explanations. Researchers first challenged a self-protection explanation by pointing out the lack of consistent support for the effect of threat on one robust and commonly used indicator of positivity bias, namely, the better-than-average effect (Chambers & Windschitl, 2004). Alternative views argued that if better-than-average judgments were best explained by a self-protection motivation, then explicit threats to the self should increase better-than-average judgments. However, a few recent studies have since demonstrated that explicit threats do elicit increased better-than-average judgments (Beer, Chester, Hughes, forthcoming; Brown, 2012; Vohs & Heatherton, 2004). While recent studies provide evidence against the first challenge to a self-protection explanation, other challenges remain.

One reason for looking beyond a strictly self-protective explanation of positivity bias is that people tend to make positively-biased self-evaluations even when threat is not heightened or made salient (Chambers & Windschitl, 2004; Klayman et al., 1999; Metcalfe, 1988; Moore & Small, 2007). For example, previous research shows that manipulating factors other than explicit threat, such as the breadth of a trait's construal,

reduces positively-biased self-evaluations (Dunning et al., 1989). Trait breadth refers to the diversity or number of behaviors that define a trait (Buss & Craik, 1983; Hampson, John, & Goldberg, 1986). For example, narrowly-construed traits such as 'tidiness' restrict the range of behaviors that can be easily associated with the self compared to broadly-construed traits such as 'talent' that have a much wider range of behaviors that can be associated with the self. People's positively-biased tendency to claim that they have more desirable personalities than their peers is attenuated when the comparisons are made for narrowly construed traits that restrict the information relevant to a judgment compared to broadly-construed traits (Dunning et al., 1989; and for evaluations of romantic partners: Neff & Karney, 2002, 2005). In addition, people tend to be overconfident about their knowledge in certain domains (e.g., temperatures in foreign cities) but not about their knowledge in other domains (e.g., poverty levels in US states) (Klayman et al., 1999). It is unlikely that a lack of knowledge of temperature information is more threatening than a lack of knowledge of poverty information (Klayman et al., 1999), which lends support to the notion that other, non-self-protective explanations may be sufficient to explain positivity biases in self-evaluation.

Second, research suggests that people do not always provide positively-biased self-evaluations. For example, people do not provide positively-biased responses in certain domains, such as evaluations of their social status (Anderson et al., 2006). In addition, people sometimes provide negatively-biased self-evaluations (Blanton et al., 2001; Chambers, Windschitl, & Suls, 2003; Kruger, 1999; Kruger & Burrus, 2004; Moore & Small, 2007). For example, people evaluate themselves to be worse than other

people at juggling, writing computer code, or coping with the death of a loved one. Moreover, people believe they are less likely than other people at living past age 100, graduating in the top 1% of their class, or owning an airplane. These negatively-biased self-evaluations share a common feature, namely, that they occur in domains in which success is rare, even though these rare abilities are no less socially desirable or important than more common abilities that tend to be characterized by positivity biases. These findings seem incongruent with a self-protection account, and raise the possibility that self-protection may not be a necessary or dominant explanation for positively-biased self-evaluation and point to alternative or parallel explanations that may sufficiently account for positively-biased self-evaluations.

Alternative views propose that aspects of the judgment stimuli and a reliance on general judgment approaches that are not tied to self-protection concerns may be sufficient causes of positivity bias (Chambers & Windschitl, 2004; Fiske & Taylor, 1991; Metcalfe, 1988; Nisbett & Ross, 1980; Pronin, Gilovich, & Ross, 2004; Ross & Sicoly, 1979; Tversky & Kahneman, 1974). In particular, alternative views address the challenges raised against a self-protection explanation by demonstrating that general judgment processes that are not tethered to self-protection motivation may lead to positively-biased responses in some situations and to a reduction in positively-biased responses or even negativity biases in other situations (Chambers & Windschitl, 2004; Klar & Giladi, 1997, 1999; Kruger, 1999; Moore & Small, 2007). While alternative views may provide a sufficient explanation for why a common mechanism elicits positivity biases and negativity biases, another possibility is that both perspectives are

correct in different motivational contexts. As mentioned above, a reliance on general judgment approaches is not a surefire way to protect the self from explicit threat because these approaches also lead to negativity bias, which would undermine a self-protection goal. Therefore, threat may engage a distinct process that is more likely to be successful at eliciting positively-biased self-evaluations.

What Can Neuroimaging Tell Us About How Positivity Biases Occur?

Why is there a gap in our knowledge about whether positivity biases reflect a single phenomenon or different phenomena as a function of whether threat is explicitly heightened? One reason is because behavioral measures alone are limited in their ability to resolve questions about the processes that underlie similar behaviors as a function of different motivations or contexts. For example, it is difficult to adjudicate with behavioral indices whether positively-biased self-evaluations reflect a self-protection motivation, a parallel or independent judgment strategy, or both, because the behavioral indices (i.e., positively-biased responses) are similar across threatening and not explicitly threatening contexts. Although behavioral research has identified a number of variables that moderate positivity biases, these moderator variables can often be explained from both a self-protection and a non-self-protection perspective. Therefore, the mechanisms that underlie positively-biased self-evaluations in different motivational contexts are underspecified.

One way to begin to understand whether positively-biased self-evaluations represent a single phenomenon or whether positively-biased self-evaluations arise from self-protective processes and more generalized judgment strategies as a function of

contextual motivators is to examine their underlying patterns of neural activation. In particular, neuroimaging methods are well suited to answer questions about how different psychological phenomena are implemented in the brain, which can inform our understanding of the processes underlying those psychological phenomena (Cacioppo & Bernston, 1992; Henson, 2006; Kosslyn, 1999; Mitchell, 2006; Ochsner & Lieberman, 2001; Posner & DiGirolamo, 2000). For example, identifying the neural systems that underlie positively-biased self-evaluations is useful because we can then examine how the engagement of those neural systems is influenced by the presence of a heightened self-protection motivation elicited by an explicit threat. If a self-protection motivation elicited by explicit threat changes the engagement of the neural systems related to positively-biased evaluations or engage additional neural regions, this might suggest that positively-biased evaluations may represent distinct phenomena as a function of contextual motivators. On the other hand, if a self-protection motivation elicited by explicit threat does not affect the neural systems associated with positively-biased evaluation, this might suggest that positivity biases represent a unitary phenomenon across different contexts. However, it can be problematic to rely too heavily on reverse inference to interpret whether the similar or distinct patterns of neural activation support similar or distinct psychological mechanisms. Therefore, identifying the precise mechanisms supported by the neural regions involved in positively-biased evaluation may provide a deeper understanding of how positively-biased evaluations are accomplished in different motivational contexts. In sum, examining positivity biases from a behavioral and neural level of analysis may have important implications for

understanding whether positivity biases reflect a self-protection motivation, a generalized judgment approach, or both. However, current neural research has not yet examined positivity biases in self-evaluation from a neural level of analysis (Beer, 2007).

Overview of the Studies

This dissertation represents a step toward answering how positivity biases are accomplished in different motivational contexts by examining the neural mechanisms that underpin them. In order to gain traction on this question, the four studies presented here all combine neuroimaging methods (fMRI) and established behavioral paradigms drawn from the social psychological literature on positivity biases in self-evaluation. The four studies take a step away from underspecified behavioral comparisons of positively-biased evaluations in different contexts to attempt to specify the mechanisms that underlie them. To this end, the specific aims of this research program are to first identify a core set of neural regions that is associated with positively-biased evaluation (Studies 1A and 1B), then build upon these initial findings to examine how the engagement of this core set of neural regions changes when self-protection is heightened by explicit threat (Study 2), and finally to examine what precise psychological processes are instantiated in the neural regions associated with positively-biased judgment (Study 3).

The first aim addresses a gap in existing neural research on self-evaluation by identifying a core set of neural regions that is associated with positively-biased jdugment. Although existing neural research on self-evaluation has identified neural regions involved in self-judgments as compared to judgments about other people and inanimate

objects (for reviews, see Amodio & Frith, 2006; Gillihan & Farah, 2005; Mitchell, 2009; Ochsner et al., 2005), it has not addressed the neural regions associated with the systematic positivity biases that sometimes affect self-evaluation and social cognition (Beer, 2007). In order to begin to address whether positivity biases represent a unitary phenomenon or multiple distinct phenomena in different motivational contexts, research is needed that first identifies a core set of neural regions associated with positively-biased evaluations. Study 1A examined the neurobiology of one robust and commonly used indicator of positivity bias, namely, the better-than-average effect. Trait breadth was manipulated in order to generate variance in positively-biased evaluations and identify neural systems associated with increased susceptibility to positivity bias. Study 1B sought to decouple better-than-average judgments from the trait breadth manipulation in order to provide converging evidence for the neural systems associated with positively-biased evaluation. For example, neural activation associated with positively-biased self-evaluation in Study 1A may be driven by susceptibility to better-than-average judgments as a function of trait breadth, but neural activation may also be driven by properties of the judgment stimuli (i.e., specificity of trait words). Therefore, Study 1B examined the neural activation associated with evaluations of social targets (close others, non-close others) that differ in their tendency to exhibit better than average responses as a function of trait breadth (Neff & Karney, 2002, 2005; Suls et al., 2002; Taylor & Koivumaki, 1976).

The second aim is to examine whether a heightened self-protection motivation changes the engagement of the neural systems associated with positively-biased self-

evaluation. Existing research has not yet examined whether explicit threat affects the neural systems associated with positively-biased self-evaluation, or whether explicit threat engages additional neural systems to compensate for threat. Therefore, it remains unknown whether positivity biases in self-evaluation when threat is explicitly heightened draws on similar or distinct neural mechanisms as positively-biased self-evaluations when threat is not explicitly heightened. Study 2 addresses this question by examining the patterns of neural activation associated with positively-biased self-evaluation elicited by an explicit threat. In Study 2, positivity bias was measured using the same approach as Studies 1A and 1B ("better-than-average" judgments) to maximize comparability between results, but elicited positivity bias in self-evaluation with an explicit threat manipulation.

The third aim is to begin to understand the psychological processes that are supported by neural regions associated with positively-biased evaluation. While the first and second aims seek to identify a set of neural regions that are associated with positively-biased evaluation and how the engagement of this set of regions is affected by self-protection motivation, they do not provide information about the precise mechanisms supported by those regions. Elucidating the mechanisms that are supported by neural regions associated with positively-biased evaluation may lead to a deeper understanding of whether positivity biases represent a single phenomenon or multiple phenomena as a function of the motivational context. Study 3 takes a step towards addressing this question by combining a signal detection approach and a contextual manipulation that permits the measurement of a process that influences the expression of positively-biased

evaluation in different contexts. Specifically, Study 3 directly tests whether neural activation associated with positivity bias tracks shifts in decision thresholds that influence the expression of positively-biased responses as a function of the motivational context.

Taken together, this dissertation has the potential to begin to uncover whether positivity biases represent a unitary phenomenon or multiple distinct phenomena by expanding our understanding of the neurobiology of the positivity biases that sometimes characterize self-evaluation. It is our goal that this line of research will motivate future work that can test how specific mechanisms may be instantiated in the neural systems associated with positivity bias.

AIM 1: WHAT NEURAL SYSTEMS ARE ASSOCIATED WITH POSITIVITY BIAS?

Study 1A

INTRODUCTION

Study 1A takes a first step towards identifying a core set of neural regions that is associated with positively-biased self-evaluations. Current neural research on self-evaluation has focused on the self-referent effect in memory by comparing neural regions that differentiate self-judgments of personality traits from judgments about personality traits of other people or inanimate objects (for reviews, see Amodio & Frith, 2006; Mitchell, 2009; Ochsner et al., 2005; Uddin et al., 2007). Taken together, these studies find that judging the personality traits of the self are robustly associated with medial prefrontal cortex (MPFC) and posterior cingulate cortex (PCC) function. However, existing neural studies do not take into account the ways in which people's self-representations tend to be positively-biased. Therefore, more research is needed to identify the neural systems that underlie positivity bias in self-evaluation.

Study 1A examined the neural systems underlying positively-biased self-evaluation by examining a robust and commonly used indicator of positivity bias, namely, the tendency for people to evaluate their personalities more favorably than the personality of their average peer (i.e., the "better-than-average" effect). Trait breadth was manipulated in order to create variance in the extent to which social comparisons are positively-biased (Dunning et al., 1989; also see Buss & Craik 1983; Hampson, John, & Goldberg, 1986). As mentioned above, the tendency to claim more desirable personalities

than the average peer is attenuated when the comparisons are made for narrowly construed traits (e.g., 'tidy') that restrict the range of behaviors that can be associated with a trait as compared to broadly construed traits (e.g., 'talented')(Dunning, Meyerowitz, & Holzberg, 1989). Therefore, neural regions that are modulated by better-than-average responses should differentiate judgments of broadly construed traits that tend to be more positively-biased from judgments of narrowly construed traits that tend to be less positively-biased. Results of this study are also reported in a published manuscript (Beer & Hughes, 2010).

METHOD

Participants

Twenty right-handed participants (9 female, 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 were compensated $15/hour or course credit for their participation. All participants were native English speakers and screened for medications or psychological and/or neurological conditions that might influence the measurement of cerebral blood flow.

Task

Participants completed a modified version of a social comparative task used in previous research (Dunning, Meyerowitz, & Holzberg, 1989). To ensure that there was a comparable "average peer" across our sample, participants were all students at the

University of Texas at Austin and judged their personality characteristics in relation to the average University of Texas student of their same gender and age (Chambers & Windschitl, 2004). In each trial, participants rated how they compared on a personality trait using a 5-point scale (-2=Much less than the average UT student; 0=About the same as the average UT student; 2=Much more than the average UT student). After each judgment, a screen depicting a fixation point indicated that participants should clear their minds (screens were jittered with lengths of 2 s (50%), 4 s (25%), or 6 s (25%) to maximize independence across experimental conditions: Donaldson, Peterson, Ollinger, & Buckner, 2001).

Participants completed 50 randomly intermixed trials of each of the Positive-Specific, Positive-Broad, Negative-Specific, and Negative-Broad conditions equally divided across 2 runs lasting 9 minutes and 10 seconds. Stimuli were projected onto a screen mounted on the bed of the scanner. Participants' head motion was limited using foam padding. Stimulus presentation and response collection was controlled by the program E-prime running on a Windows 98 Computer.

Stimuli

Trait words were equally distributed across Valence and trait Breadth. Stimuli were selected from trait word lists which have been standardized for valence, breadth, familiarity, and number of syllables (Anderson, 1968; Kirby & Gardener, 1972) and used in many previous behavioral and neural studies of self-processing (Alicke, 1985; Dunning, Meyerowitz, & Holzberg, 1989; Kelley et al., 2002; Moran et al., 2006;

Ochsner et al., 2005). To ensure that this information was not outdated, a sample of 10 student judges who would be representative of our fMRI study population rated 250 words for valence (i.e., social desirability), trait breadth, familiarity, and judgment certainty. Ratings were consistent with standardized information (Anderson, 1968; Kirby and Gardener, 1972). The 200 words used for the experiments were selected using several constraints. Words that were not familiar to at least one of our judges were eliminated. Four sets of 50 words based on the published norms and our student judges were equated for (a) social desirability within valence level (e.g., positivity of positive-broad vs positive-specific traits, negativity of negative-broad vs negative-specific traits, $p > .05$) and (b) judgment certainty ($p > .05$) but (c) differed in trait breadth (positive-broad vs positive-specific, $t = 11.4$, $p < .05$; negative-broad vs negative-specific, $t = 26.7$, $p < .05$). These criteria ensured that traits differed in their breadth but not in additional factors such as familiarity, social desirability, or self-descriptiveness (measured by certainty: Sedikides, 1993; 1995).

The Positive-Specific condition consisted of trait words such as prompt, talkative, tactful, coolheaded, mathematical, well spoken, witty, modest, energetic, and lighthearted. The Positive-Broad condition consisted of trait words such as likable, mature, decent, positive, capable, understanding, educated, competent, disciplined, and ethical. The Negative-Specific condition consisted of trait words such as stingy, materialistic, bashful, high strung, rigid, gullible, timid, jumpy, boastful, and messy. The Negative-Broad condition consisted of trait words such as lacking, bad, weak, maladjusted, irritating, unreliable, phony, narrow minded, aggressive, and showy.

FMRI 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. Functional images were acquired with a GRAPPA sequence (TR = 2000ms, TE = 30 ms, FOV=240, voxel size 2.5 mm x 2.5 mm x 3 mm) with each volume consisting of 35 axial slices in line with the AC-PC line. These parameters were implemented to optimize coverage of the orbitofrontal cortex without sacrificing whole-brain acquisition. A high resolution SPGR T1-weighted image was also acquired from each subject so that functional data could be normalized to the Montreal Neurological Institute (MNI) atlas space.

FMRI Data Analysis

All 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. Image volumes were corrected for slice-timing skew using temporal sinc-interpolation and for movement using rigid-body transformation parameters. Structural and functional volumes were normalized to T1 and EPI templates, respectively, using a 12-parameter affine transformation together with a nonlinear transformation involving cosine basis functions that resampled the volumes to 2-mm cubic voxels. Images were then smoothed with an 8-mm FWHM Gaussian kernel. To remove drifts within sessions, a high-pass filter with a cutoff period of 128 seconds was applied.

A fixed-effects analysis modeled event-related responses for each participant. Responses related to judgment in the Positive-Specific, Positive-Broad, Negative-Specific, and Negative-Broad conditions were modeled as events using a canonical hemodynamic response function with a temporal derivative. A general linear model analysis created contrast images for each participant summarizing differences of interest. Contrasts from each participant were used in a second-level analysis treating participants as a random effect. Group average SPM{t} maps were created for contrasts of interest (Specific > Broad; Broad > Specific).

Interpretation of results from main contrasts was limited to regions that had previously been associated with self-referential processing, valence, availability heuristics, and emotional reappraisal (e.g., Beer, in press; Ochsner et al., 2005; DeMartino et al., 2006; Krusemark, Campbell & Clementz, 2008; Moran et al., 2006; Sharot et al., 2007a). Contrasts of interest were masked by a priori neuroanatomical VOIs from the Automated Anatomical Labeling map (Tzourio-Mazoyer et al., 2002) and activation clusters that survived correction for multiple comparisons ($P < .05$ familywise error (FWE), k=10) were reported (search volumes: lateral orbitofrontal cortex (LOFC: 23-mm^3), medial orbitofrontal cortex (MOFC: 17-mm^3), MPFC (19-mm^3), ventral anterior cingulate cortex (vACC: 22- mm^3), dorsal anterior cingulate cortex (dACC: 22-mm^3), posterior cingulate cortex (PCC: 15-mm^3), and insula (25-mm^3). Parameter estimates (i.e., beta weights) were extracted from significant clusters using Marsbar (Brett et al., 2002). The parameter estimates represent the regression coefficients from the main contrasts in the general linear model predicting MR signal.

Multiple regression tested whether individual differences in ratings in the Specific and Broad trait conditions predicted neural differentiation in the Specific > Broad contrast (ratings Positive Specific and Positive Broad, $r = .81$, $p < .05$; ratings Negative Specific and Negative Broad, $r = .86$, $p < .05$). Ratings for Negative traits were reverse-scored so they could be collapsed with ratings of Positive traits to reflect average deviation from the average peer. Results from the regression analyses were corrected for multiple comparison at $p < .05$ FWE based on the activation clusters from the group contrasts of the Specific > Broad contrast (8-mm^3 volume around main effect peaks).

RESULTS

Task Performance

No gender differences were found in responses or reaction times ($Fs < 1$) so results are reported collapsed across gender. Consistent with previous behavioral research (Dunning, Meyerowitz, and Holzberg, 1989), self-evaluations were characterized by a significant interaction between the Valence (Positive, Negative) and Breadth (Broad, Specific) factors ($F(1, 19) = 108.75$, $p < .05$; Figure 1) that qualified a main effect of Valence ($F(1, 19) = 74.80$, $p < .05$). In comparison to the average peer, participants on average viewed themselves as significantly more likely to have the Positive-Broad traits ($t(19) = 5.45$, $p < .05$) and significantly less likely to have Negative-Broad traits ($t(19) = -7.94$, $p < .05$) when compared to their respective Specific conditions. Participants did not just claim positive traits and downplay negative traits; they tended to view themselves as most distinct for positive and negative words for traits with broader construals.

Participants' reaction times were characterized by main effects of Valence ($F(1,19) = 15.4$, $p < .05$) and Breadth ($F(1,19) = 10.0$, $p < .05$) but their interaction did not reach significance ($F(1,19) = 2.6$, $p > .05$). Judgments in the Positive condition (M = 1426.62 ms, SE = 25.9) were made more quickly than judgments in the Negative condition (M = 1466.21 ms, SE = 22.8). Judgments in the Broad condition (M = 1432.58 ms, SE = 25.6) were made more quickly than judgments in the Specific condition (M = 1460.25 ms, SE = 23.1).

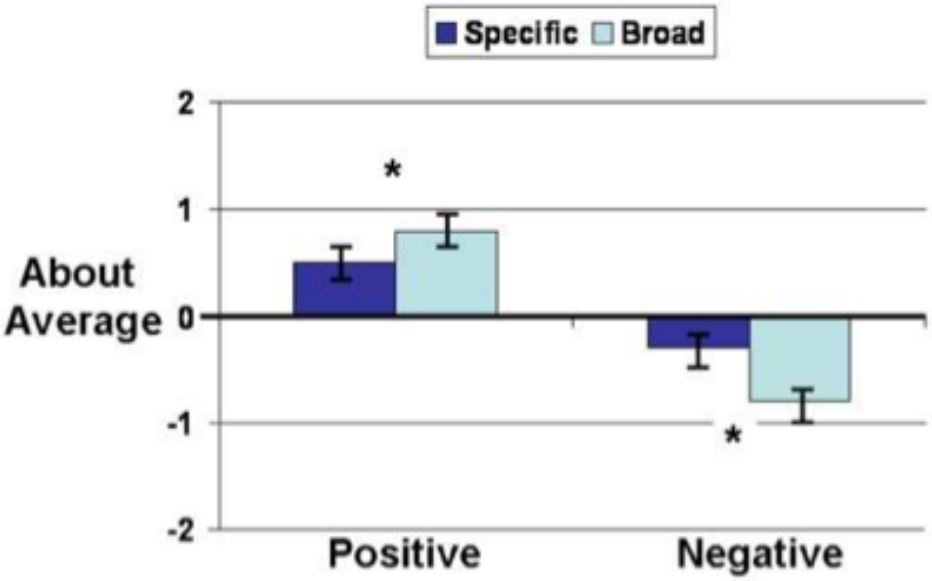

Figure 1: Behavioral results for social comparison evaluations for self

Means and standard errors of Self-evaluations of positive and negative traits in comparison to an average peer. On average, the sample should estimate their traits at the midpoint of the scale (''0'') for unbiased evaluations.

FMRI Results

MPFC, OFC, and dACC are Associated with Social-Comparative Judgments of Specific vs Broad Traits

A direct comparison between the Specific condition and the Broad condition was used to examine neural regions associated with susceptibility to "above average" judgments. In contrast to the Broad trait condition, judgments of Specific traits were associated with greater activation in MPFC (peaks = Brodmann's Area (BA) 10: 8, 64, 24, and BA 9 = 12, 54, 34). Additionally, the Specific > Broad contrast revealed significant activation in medial OFC (BA 11 peaks = -4, 46, -10, and -2, 56, -14), lateral OFC (left BA 47 peak = -32, 34, -14, right BA 47 peak = 28, 28, -20) and dACC (BA 24 peak = 10, 26, 34)(See Figure 2A-B). No significant activation was found for the main contrast of Broad > Specific.

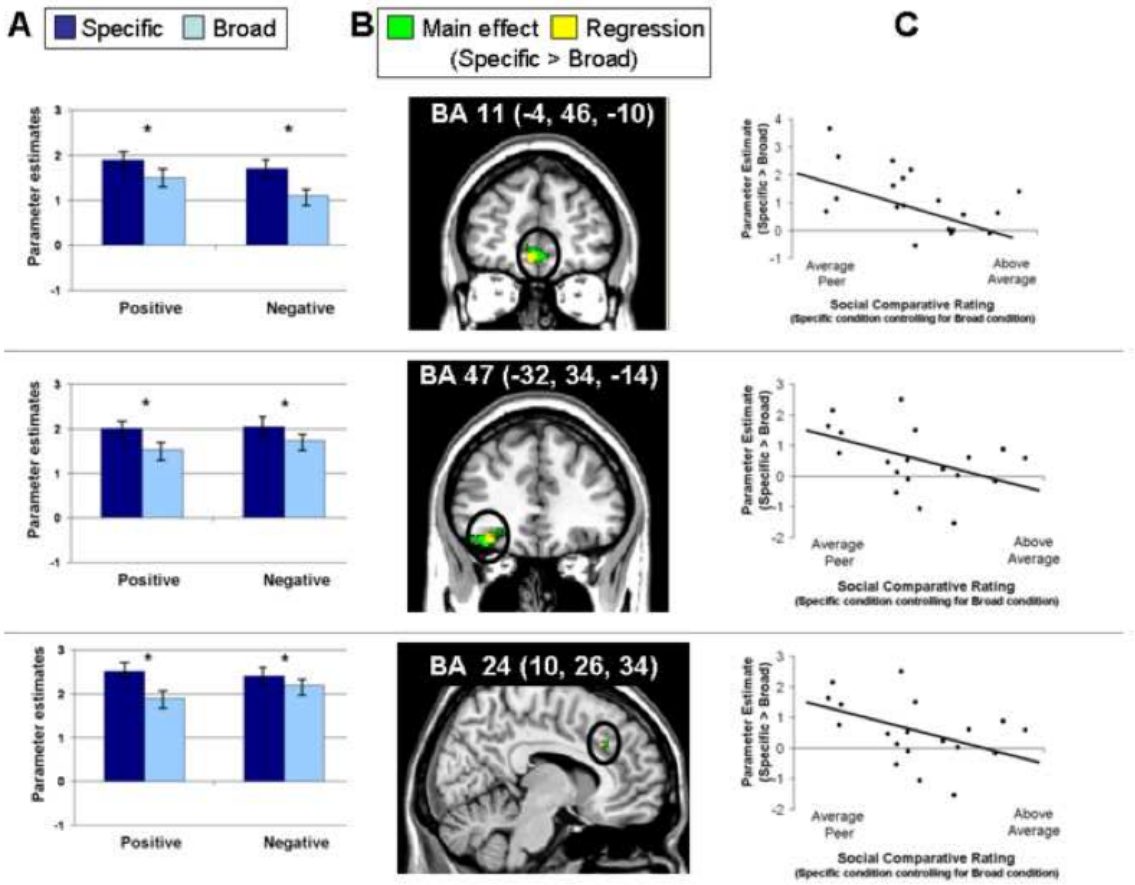

Figure 2: Neural regions associated with reduced better-than-average responses for self

A: Parameter estimates for the Specific > Broad contrast. "*" indicate significant differences. B: Neural activation associated with main effect and the overlap of the regression analysis for Specific > Broad contrast: mOFC: y =46; lOFC: y = 34; dACC: x = 10. C: Scatter plots depict the regression analysis of individual differences in social-comparative ratings on neural activation.

OFC and dACC are Negatively Modulated by Individual Differences in "Better Than Average" Judgments

Neural regions that differentiate the Specific and Broad conditions have two possible interpretations: they might be related to the differences in social-comparative

judgments or they might be related to the fact that participants were processing relatively specific or broad traits irrespective of social-comparative judgments. Therefore, we conducted a regression analyses to examine whether the neural activation associated with the Specific > Broad contrast was driven by individuals who tended to rate themselves as more similar to the average peer. The more participants viewed themselves like their average peer in the Specific condition, the more they recruited regions of the OFC and dACC activation identified in the main contrast of Specific > Broad (see Figure 2B-C). Individual differences in "better than average" judgments were negatively correlated with activation in medial OFC (regression peak = -6, 46, -10; $t = 2.96$, $p < .05$ FWE) and left lateral OFC (regression peak = -34, 34, -16; $t = 2.83$, $p < .05$ FWE). Similarly, there was a trend for negative correlation between individual differences in "above average" judgments and dACC activation (regression peak = 10, 26, 32; $t = 2.51$, $p = .07$ FWE).

DISCUSSION

Study 1A represents a first step toward identifying a core set of neural regions that is associated with positively-biased self-evaluation. While existing neural research finds a robust association between MPFC function and self-evaluation, it does not address the neural systems that are associated with the tendency for self-evaluations to be positively-biased. The current study builds on previous neural research examining self-judgments of personality traits by examining the tendency for people to evaluate their personality traits more favorably in comparison to their average peer. Consistent with previous behavioral research, participants evaluated their personality traits as better than the personality of

their average peer; this tendency was reduced for evaluations of specific traits that are characterized by a more restricted range of behaviors that are associated with a trait as compared to broad traits. The MPFC, a region often associated with self-evaluation, showed significantly increased activation for judgments of specific traits as compared to broad traits. However, the MPFC region was not modulated by individual differences in better-than-average responses. OFC and dACC were also significantly more activated for judgments of specific traits compared to broad traits. Unlike the MPFC, activity in medial and lateral OFC and, to a lesser extent, dACC was negatively modulated by individual differences in better-than-average responses. The more participants recruited OFC and, to a lesser extent, dACC activation, the less they evaluated themselves as better than their average peer. The present findings provide new evidence for a core set of neural regions associated with positivity biases in self-evaluation.

The findings extend and contribute to a large body of research on the neural systems involved in self-evaluation. As mentioned above, current neural research on self-evaluation have identified MPFC and PCC as important regions for evaluating the self as compared to non-social stimuli (e.g., Amodio & Frith, 2006; Mitchell et al., 2006; Ochsner et al., 2005; Uddin et al., 2007), but existing research had not yet considered the biases that sometimes pervade self-evaluation. The findings from Study 1A suggest that discussions about the neural systems of self-evaluation should be expanded to include a role for OFC and dACC in biased self-evaluation.

While the results from the present study suggest that reduced OFC and dACC activation may be associated with positively-biased evaluations, more research is needed

to more deeply understand the set of neural regions associated with positively-biased evaluation. In the present study, trait breadth was manipulated in order to create variance in the extent to which evaluations were positively-biased. The association between OFC and dACC function and reduced positivity bias was drawn from a contrast of neural activity during the specific trait condition associated with reduced positivity bias compared to neural activity in the broad trait condition associated with positivity bias. Therefore, OFC and dACC activation may be associated with positivity bias, but OFC and dACC activation may be related to other factors, such as differences in the stimulus properties themselves (e.g., specificity of the trait words). The individual difference analysis provides evidence against this interpretation, suggesting that the most likely explanation is that OFC and dACC activation are involved in reduced better-than-average ratings. In addition, the present results are consistent with research showing that patients with OFC damage tend to evaluate their social behavior more favorably than the evaluations of trained judges (Beer et al., 2006). Additionally, source localization analyses from an ERP study suggest that dACC activation may be associated with non-self-serving attributions of success on a working memory task (Krusemark, Campbell, & Clementz, 2008), Lastly, OFC has been shown to attenuate biases in non-social judgments (De Martino et al., 2006). However, research that decouples better-than-average responses from the trait breadth manipulation is needed to more deeply understand the neural association of positivity bias.

Study 1B

INTRODUCTION

The aim of Study 1B is to provide converging evidence for the neural systems associated with positively-biased evaluation. Study 1A suggested an association between OFC and dACC activation and reduced better-than-average responses as a function of trait breadth (Beer & Hughes 2010). However, OFC and dACC activation may have also been related to properties of the judgment stimuli (i.e., specificity of trait words). Therefore, the goal of Study 1B is to decouple better-than-average responses from the trait breadth manipulation in order to better understand the neural association of positively-biased evaluation.

To this end, Study 1B examines the neural systems associated with evaluations of social targets (Close Other, Non-Close Other) that differ in their susceptibility to better-than-average responses elicited by trait breadth (Neff & Karney, 2002, 2005; Suls et al., 2002). Previous research has shown that evaluations of Close Others tend to be better-than-average for broad traits, whereas evaluations of Close Others for specific traits and evaluations of Non-Close Others for *all* traits tend to be more similar to the average peer (Neff & Karney, 2002, 2005; Suls et al., 2002). Therefore, if OFC and dACC activation are associated with reduced positively-biased social comparisons and not to aspects of the judgment stimuli, then OFC and dACC activation should be greater for judgments of Close Other's Specific traits and *all* Non-Close Other's traits, as compared to Close Other's Broad traits. Finally, individual differences in the degree to which people

evaluate their Close Others and Non-Close Others as better-than-average should be negatively correlated with OFC and dACC activity. However, if OFC and dACC are related to properties of the judgment stimuli, then these regions should also differentiate Specific from Broad traits for Non-Close Others despite no differences in better-than-average judgments between Non-Close Other Specific and Broad traits. Results of this study are also reported in a published manuscript (Hughes & Beer, in press-a).

METHOD

Participants

Twenty right-handed participants (15 females, M age = 18.7 years, SD = 0.8 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. In addition, all participants were prescreened to ensure that each had a romantic partner and a roommate. Participants with more than 1 roommate were instructed to select one of them for the purpose of the study. Participants whose roommates were biologically related to them (i.e., siblings, cousins, etc.) were excluded from participation.

Task

Participants completed a modified version of a social comparison task used in Study 1A and previous research (Dunning et al., 1989; Beer & Hughes, 2010). In the

task, participants compared the personality traits of a Close Other (i.e., their romantic partner) and a Non-Close Other (i.e., their roommate) with the personality traits of an average peer of their same age and gender at their university (see Figure 3). Comparisons were made in relation to an average peer of the same age and gender as the Close Other and Non-Close Other to ensure that there was a comparable ‘‘average peer’’ across our sample. Participants were presented with personality trait words and had to make comparisons for Close Others and Non-Close Others (see Figure 3) using a 5-point scale (–2 = much less than the average UT student; –1 = slightly less than the average UT student; 0 = about the same as the average UT student; 1 = slightly more than the average UT student; 2 = much more than the average UT student).

Following previous research on social cognition (e.g., Kelley et al. 2002; Ochsner et al. 2005), we used 2 different cues to remind participants which Target (Close Other, Non-Close Other) they were comparing with an average peer. First, participants were presented with a 2 s instruction screen that indicated the Target for comparison (Close Other, Non-Close Other) (see Figure 3). Second, each instruction screen was followed by a set of probes from the social comparison task for that Target. Each probe reminded the participants what Target was of interest and indicated the personality trait word of interest (see Figure 3). Within a set of probes, personality trait words were 1) randomly sampled from 4 trait categories described below (see Stimuli) and 2) jittered with screens depicting a fixation point. Participants were instructed to clear their minds when they saw a screen with a fixation point. These fixation screens were randomly jittered (2 s [50%], 4 s [25%], 6 s [25%]) to maximize independence across experimental (Donaldson et al.,

2001). This approach provides strong reminders of the Target of interest and allows independent modeling of the neural activation for each social comparison rating.

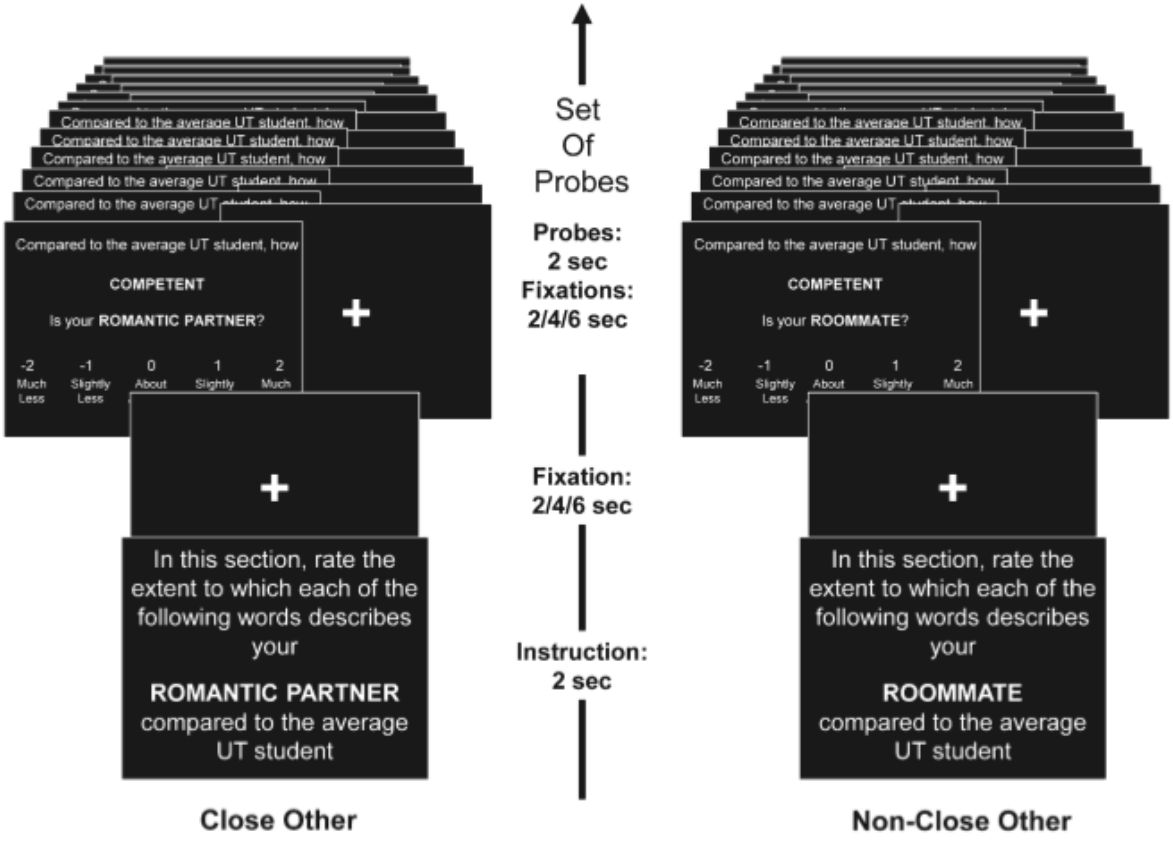

Figure 3: Stimuli and timing in social-comparative judgment task

Participants saw an instruction screen that indicated the Target (‘‘Romantic Partner’’ or ‘‘Roommate’’) of judgment. The instruction screen was separated from a subsequent set of social comparison probes with a jittered screen depicting a fixation point. Within a set of probes, personality trait words were jittered with screens depicting a fixation point.

FMRI data were collected while participants performed 4 functional runs of the social comparison task described above. Each functional run lasted 10 min 4 s. Within a run, participants rated 4 sets of probes for each of the Close Other and Non-Close Other condition. The presentation order of the probe sets was randomly assigned within a run.

In order to present the 200 traits words (50 words each for 4 categories, see Stimuli below) for each Target across 4 runs which were each divided into 4 sets of probes, it was necessary to have 50% of the probe sets include 12 probes and the other 50% include 13 probes (e.g., 200/16 = 12.5). Stimuli were projected onto a screen mounted on the bed of the scanner and head motion was limited using foam padding. Stimulus presentation and response collection was controlled by the program E-prime running on a Windows XP Computer.

After leaving the scanner, participants rated each Target on Duration of Relationship (''How long have you known your roommate/romantic partner (in months)?''), Liking (''How much do you like your roommate/romantic partner?''), Similarity (''How similar do you consider yourself to your roommate/romantic partner?''), and Close- ness (''How close are you to your roommate/romantic partner?'') on a 5- point scale (descriptive anchors were provided for 3 points: 1 = not very much; 3 = somewhat; 5 = very much; points 2 and 4 reflected the continuum between the extreme endpoints and midpoint). Liking, Similarity, and Closeness for Non-Close Others were all highly correlated (all r s > 0.65, P < 0.05). Therefore, a composite score ''Intimacy'' was created from these variables. There was a ceiling effect for Liking and Closeness ratings for Close Others, so the composite score Intimacy was not created for Close Others. Participants also rated themselves on the social comparison task (i.e., compared themselves with the average peer).

FMRI 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. Functional images were acquired with a GRAPPA sequence (time repetition = 2000 ms, time echo = 30 ms, field of view = 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 OFC without sacrificing whole-brain acquisition. A high-resolution SPGR T1-weighted image was also acquired from each subject.

FMRI Data Analysis

All statistical analyses were conducted using SPM2 (Wellcome Department of Cognitive Neurology). Functional images were recontructed from k-space using a linear time interpolation algorithm to double the effective sampling rate. Image volumes were corrected for slice-timing skew using temporal sinc interpolation and for movement using rigid-body transformation parameters. Functional data and structural data were coregistered and normalized into a standard anatomical space (2-mm isotropic voxels) based on the echo planar imaging and T1 templates (Montreal Neurological Institute), respectively. Images were smoothed with an 8-mm full-width at half- maximum Gaussian kernel. To remove drifts within sessions, a high- pass filter with a cutoff period of 128 s was applied.

A fixed-effects analysis modeled event-related responses for each participant. For each Target (Close Other, Non-Close Other), the Positive-Specific, Positive-Broad,

Negative-Specific, and Negative-Broad conditions were modeled as events using a canonical hemodynamic response function with a temporal derivative. A general linear model analysis created contrast images for each participant. Contrasts relevant to the hypotheses were calculated. First, contrast images were calculated to examine the interaction of trait Breadth (Specific, Broad) and Target (Close Other, Non-Close Other) collapsed across Valence (Positive, Negative) on neural activation. Based on previous research, judgments of Close Other Broad traits are more likely to be better than average as compared with judgments of Close Other Specific and all Non-Close Other traits (Neff & Karney 2002; Suls et al. 2002). Therefore, the Target X Breadth interaction contrast was modeled as (Close Other Broad –3; Close Other Specific +1; Non-Close Other Specific +1; Non-Close Other Broad +1). The Target X Breadth interaction contrast introduces the potential confound of Non-Close Other Intimacy. Previous research suggests that the motivation to cast other people (e.g., roommates) in a positive light may vary to the extent that they are more intimate and well liked (Taylor & Koivumaki 1976; Suls et al. 2002). For the minority of participants who rated their roommates high on intimacy, the neural hypotheses for the Non-Close Other (i.e., roommate) condition would more closely resemble the hypotheses for the Close Other condition. More specifically, neural regions that differentiate Broad from Specific traits for Close Others may also differentiate Broad from Specific traits for Non-Close Others (i.e., roommates) to the extent that they are more intimate and well liked. Therefore, the Target X Breadth interaction contrast controlled for Non-Close Other Intimacy once they were entered into group level analyses (see below).

Contrasts from each participant were used in a second-level analysis treating participants as a random effect. The group average SPM{t} maps were masked by a priori regions of interest (ROIs) and only clusters that survived correction for multiple comparisons ($P < 0.05$ FWE, $k = 10$) in a priori ROIs were interpreted. The ROIs were based on the activations found in Study 1A (Beer & Hughes 2010) and were defined by 8-mm-radius spheres around the peaks of activation clusters: MPFC (BA 9: 12, 54, 34; BA 10: 8, 64, 24), medial OFC (MOFC) (BA 11: –2, 56, –16 and –4, 46, –10), bilateral lateral OFC (LOFC) (left BA 47: –32, 34, –14; right BA 47: 28, 28, –20), and dACC (10, 26, 34). Parameter estimates from significantly activated clusters from relevant contrasts were extracted using Marsbar (Brett et al. 2002). The parameter estimates were then used to test for significant correlations between brain activation identified by our main contrasts and individual differences in behavioral ratings (Kriegeskorte et al. 2009; Poldrack & Mumford 2009; Vul et al. 2009).

A test for significant correlation examined whether reduced neural activation for the Close Other Broad condition as compared with the other 3 conditions was driven by individuals who also tended to rate Close Other Broad traits as more above average compared with their ratings of the other 3 conditions. Therefore, parameter estimates from significantly activated clusters from the Target X Breadth interaction contrast were tested for significant correlation with individual differences in Close Other Broad ratings compared with the Close Other Specific and all Non-Close Other ratings (i.e., Differences in Social Comparison Ratings). First, judgments of Negative traits were reverse scored so that ratings could be collapsed across valence to reflect deviation from

the average peer for the Specific and Broad trait conditions. When reverse scoring is applied to social comparisons of negative traits, higher values indicate greater above average ratings (e.g., more positive traits, fewer negative traits), whereas values closer to zero indicate greater similarity to the average peer. Individual differences in Close Other Broad ratings compared with ratings of the other 3 conditions was calculated by applying the same weights of the Target X Breadth interaction contrast to the behavioral ratings (Close Other Broad Rating –3; Close Other Specific Rating +1; Non-Close Other Specific Rating +1; Non-Close Other Broad Rating +1). With this coding scheme, high scores were closer to zero and indicated that Close Other Broad trait ratings did not differ from ratings of the other 3 conditions. On the low end, scores tended to be more negative and indicated greater above average ratings in the Close Other Broad condition compared with the other 3 conditions. Therefore, a negative correlation between this behavioral index and neural activation indicates less activation in relation to above average ratings.

RESULTS

Differences between Close Other and Non-Close Other Relationships

As a manipulation check, individual differences in Duration of Relationship, Liking, Closeness, and Similarity of Close Others were compared with Non-Close others. Duration of Relationship was significantly longer for Close Others than Non-Close Others (Close Other: $M = 37.89$ months, $SD = 23.33$; Non-Close Other: $M = 20.00$ months, $SD = 33.93$; $t(19) = 2.23$, $p < 0.05$). Close Others were rated more highly than Non-Close Others on Liking (Close Other: $M = 5.0$, $SD = 0$; Non-Close Other: $M = 3.7$,

SD = 1.3; $t(19) = 4.47$, $p < 0.05$), Closeness (Close Other: M = 5.0, SD = 0; Non-Close Other: M = 3.2, SD = 1.4; $t(19) = 5.28$, $p < 0.05$), and Similarity (Close Other: M = 4.1, SD = 0.9; Non-Close Other: M = 2.7, SD = 1.5; $t(19) = 3.56$, $p < 0.05$).

Task Performance

Consistent with previous research, no gender differences were found in responses or reaction times (F s < 1), so all results are reported collapsed across gender (Dunning et al. 1989; Kenny & Acitelli 2001; Beer & Hughes 2010). Social comparisons were characterized by a significant interaction between Valence (Positive, Negative), Breadth (Specific, Broad), and Target (Close Other, Non-Close Other, Self: 3-way interaction: $F(1,19) = 6.39$, $p < 0.05$; see Figure 4). As expected, judgments in the Broad condition were associated with greater deviations from about average in the Close Other and the post-scan Self-condition but not in the Non-Close Other condition. Both Close Others and Self were judged as significantly more likely to have Positive-Broad traits (Close Other: $t(19) = 5.70$, $P < 0.05$; Self: $t(19) = 7.36$, $p < 0.05$) and significantly less likely to have Negative-Broad traits (Close Other: $t(19) = –9.89$, $p < 0.05$; Self: $t(19) = –8.89$, $p < 0.05$) when compared with their respective Specific conditions. Participants did not just claim positive traits and downplay negative traits for their Close Others and the Self; they were most likely to exhibit better than average judgments in relation to broad traits.

In contrast, judgments of Non-Close Others were significantly differentiated by Valence ($t(19) = 2.61$, $p < 0.05$) but were not significantly differentiated by Breadth (see Figure 4). Ratings either did not significantly differ from the about average point on the

scale (e.g., Positive-Broad: $t(19) = 1.81$, $p > 0.05$; Positive-Specific: $t(19) = 0.78$, $p > 0.05$) or fell within the same range as the Specific ratings for the other targets (all $ts < 1$).

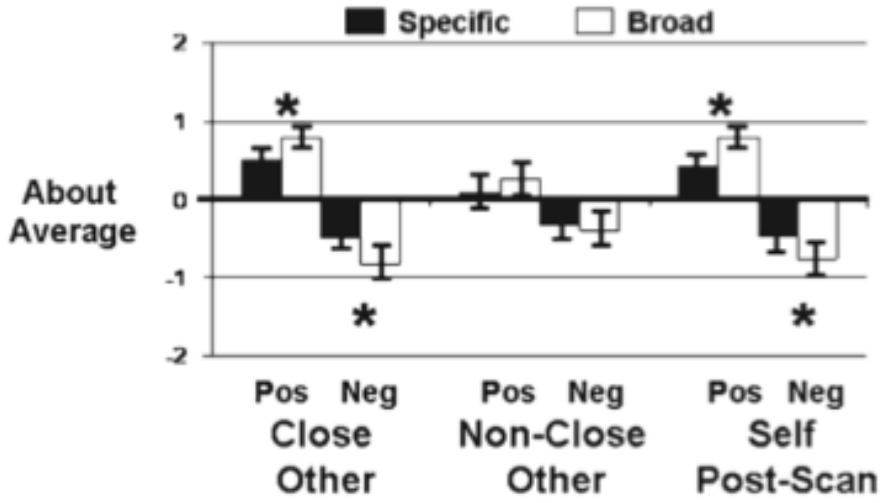

Figure 4: Behavioral results for social comparison evaluations of other people

Means and standard errors of Close Other, Non-Close Other, and (post-scan) Self-evaluations of positive and negative traits in comparison to an average peer. On average, the sample should estimate their traits at the midpoint of the scale (''0'') for unbiased evaluations.

Positive ratings were faster than Negative ratings in the Close Other and Non-Close Other conditions ($F(1,19) = 4.93$, $p < 0.05$). No significant effects were found for Breadth ($F(1,19) = 1.13$, $p > 0.05$), Target ($F(1,19) = 0.13$, $p > 0.05$), or any pairwise interaction of these variables ($Fs < 1.5$). The 3-way interaction was marginally significant ($F(1,19) = 3.73$, $p = 0.07$); this effect was driven by the especially fast reaction times for the Positive-Broad ratings of Close Other.

FMRI Results

OFC and, to a Lesser Extent, dACC Are Associated with Judgments that Are Closer to Average

No significant activation was found for MPFC or PCC in the Target X Breadth interaction contrast. Instead, the neural regions that differentiated Close Other Broad trait judgments from Close Other Specific and all Non-Close Other trait judgments (Figure 5) were the same as those associated with differentiating Broad from Specific trait judgments for the self in Study 1A (Beer & Hughes 2010). The Target X Breadth interaction contrast showed significant activation in the 1) MOFC (BA 11: peak = –10, 48, –14; t-stat = 4.08, k = 129, $P < 0.05$ FWE), 2) left LOFC (BA 47: peak = –24, 42, –14; t-stat = 4.21, k = 77, $P < 0.05$ FWE), and marginally significant activation in the 3) dACC (BA 24: peak = 14,28,30; t-stat = 3.16, k = 75, $P = 0.08$ FWE). MOFC, LOFC, and, to a lesser extent, dACC, were associated with reduced activity in the Close Other Broad condition compared with the Close Other Specific and all Non-Close Other conditions.

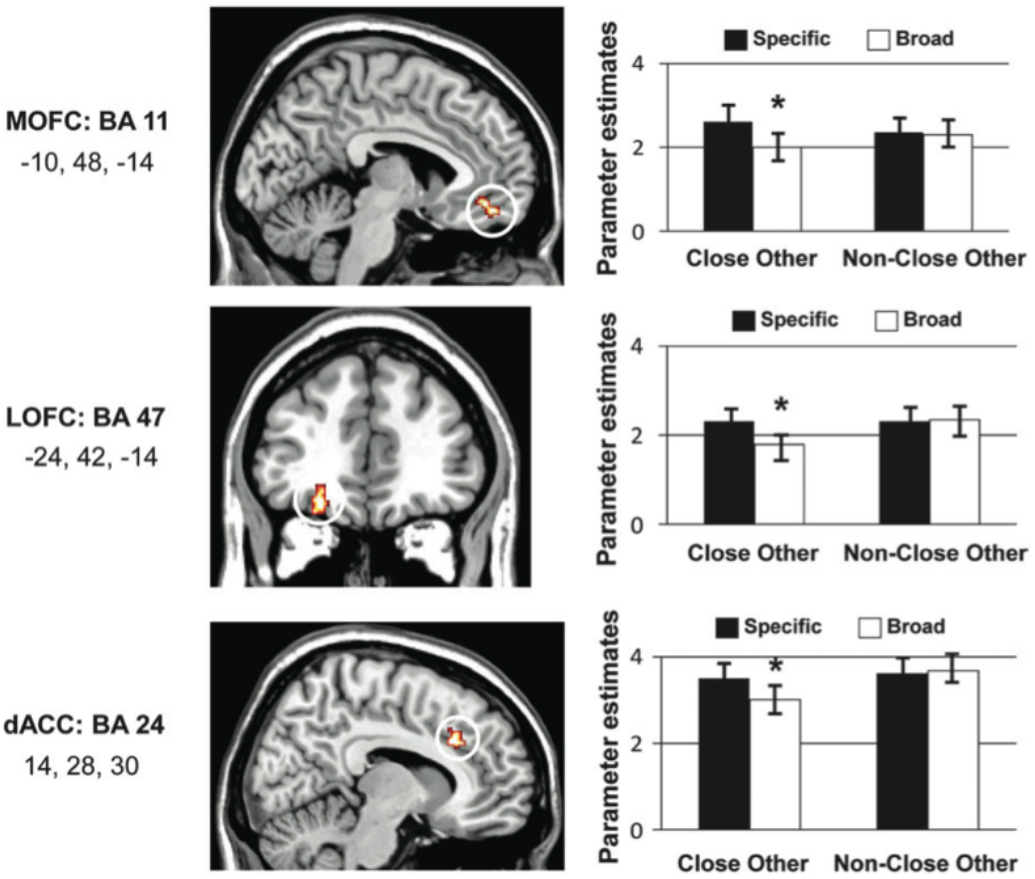

Figure 5: Neural regions associated with reduced better-than-average evaluations of other people.

Neural regions identified by the Target X Breadth interaction contrast and parameter estimates in relation to baseline extracted for each condition of Close Other and Non-Close Other (x=10: MOFC [BA 11]; y=42: LOFC [BA 47]; x=14: dACC [BA 24]). MOFC, LOFC, and, to a lesser extent, dACC, are associated with reduced activity in the Close Other Broad condition as compared with Close Other Specific and all Non-Close Other conditions.

OFC Is Negatively Modulated by Individual Differences in ‘‘Above Average’’ Judgments

The reduced neural activation for the Close Other Broad condition compared with the other 3 conditions was driven by individuals who tended to rate Close Other Broad

traits as above average compared with their ratings of the other 3 conditions. Parameter estimates from the Target X Breadth interaction contrasts were negatively correlated with behavioral indices of how much participants evaluated their close others as above average on broad traits compared with all other conditions. The more participants viewed their Close Others as better than the average peer in the Broad trait condition (compared with the other 3 conditions), the less they recruited MOFC ($r = –0.56$, $p < 0.05$) and LOFC ($r = –0.45$, $p < 0.05$) (see Figure 6).

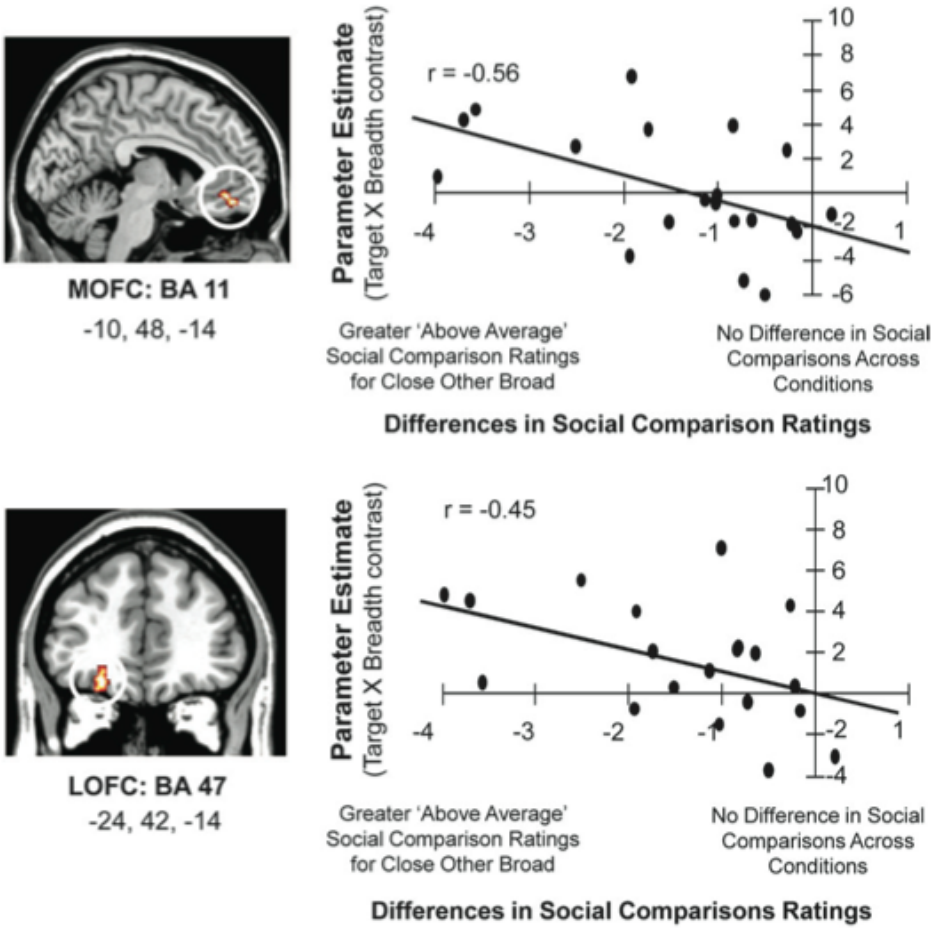

Figure 6. Individual differences in better-than-average responses modulate OFC activation

Social-comparative ratings modulate neural activation from the Target 3 Breadth interaction contrast. Scatter plots depict the correlation between individual differences in social comparison ratings and MOFC and LOFC activation identified in the Target 3 Breadth interaction contrast. The more participants rated their Close Others as better than average in the Broad trait condition compared with the other 3 conditions, the less they recruited OFC regions in the Close Other Broad condition compared with the other 3 conditions.

DISCUSSION

Study 1B provides converging evidence for the neural systems associated with positively-biased evaluation by decoupling better-than-average responses from the trait breadth manipulation. Study 1B results show that OFC and dACC activation associated with better-than-average judgments cannot be explained by the specificity of the trait stimuli. OFC and, to a lesser extent, dACC were associated with conditions of reduced better-than-average judgments, which included the Specific trait condition for Close Others, and both trait breadth conditions for Non-Close Others (i.e., Broad and Specific traits). Furthermore, the more participants viewed their Close Others and Non-Close Others as better than their average peer, the less they recruited medial and lateral OFC activation. The pattern of neural activation found in the main contrast and individual difference analysis reflect the Target (Close Other, Non-Close Other) by Trait Breadth (Specific, Broad) interaction found in better-than-average judgments rather than a main effect of Trait Breadth regardless of Social Target. If OFC and dACC activation were driven by properties of the judgment stimuli and not positively-biased social comparisons, then OFC and dACC activation should have differentiated Specific from Broad traits for Close Others *and* Non-Close Others. Instead, the present findings show

that OFC and dACC activation differentiate judgments that are better than average from judgments that are more similar to average.

The findings from Study 1B also contribute to a large body of research on the neural systems involved in social cognition. Previous research on social cognition suggests that evaluations of other people, much like self-evaluations, recruit increased MPFC and PCC activation compared to non-social judgments (e.g., Ochsner et al., 2005; Jenkins et al., 2008). Just as previous research suggests that MPFC and PCC may support processes that are engaged when evaluating the self and other people (Jenkins et al., 2008; Krienen et al., 2010; Mitchell et al., 2006; Ochsner et al., 2005), the present findings suggest that OFC and dACC may support processes that are engaged when making positively-biased evaluations about the self and other people. Taken together, Studies 1A and 1B identify a core set of neural regions, comprised of OFC and dACC function, that is modulated by positively-biased evaluation. However, Studies 1A and 1B do not address how an explicitly manipulated self-protection motivation may change the engagement of this core set of neural regions that underlie positivity bias.

AIM 2: DOES SELF-PROTECTION MOTIVATION CHANGE THE ENGAGEMENT OF NEURAL SYSTEMS OF POSITIVITY BIAS?

Study 2

INTRODUCTION

Study 2 seeks to examine whether a heightened self-protection motivation changes the engagement of the core set of neural regions associated with positivity bias, or whether additional neural associations of positivity bias are brought online. Studies 1A and 1B found that a core set of neural regions comprised of OFC and, to a lesser extent dACC was associated with reduced positively-biased self-evaluations. However, these previous studies did not elicit positivity bias by explicitly manipulating threat, so it is not known if a heightened self-protection motivation elicits positivity bias via the same neural processes or whether heightened self-protection motivation changes the engagement of those neural processes. In order to begin to understand whether positivity biases represents a single phenomenon or distinct phenomena as a function of whether self-protection motivation is heightened, research is needed that examines the effect of explicit threat on the neural associations of positivity bias identified in Studies 1A and 1B.

Study 2 takes a step toward addressing this open question by examining the neural systems associated with positively-biased self-evaluations elicited by explicit threat. As mentioned above, explicit threat increases the tendency to evaluate the self as better than the average peer presumably as a way to protect the self (e.g., Brown, 2012; Vohs & Heatherton, 2004). Therefore, a threat manipulation was used to elicit positivity bias in a

social comparison task in which trait breadth had previously been used to examine the neural systems associated with positively-biased evaluations (Study 1A: Beer & Hughes, 2010; Study 1B: Hughes & Beer, in press-a). Examining whether heightened self-protection motivation changes the engagement of the neural systems associated with positivity bias in self-evaluations may begin to inform our understanding of whether positivity biases represent a single phenomenon or multiple distinct phenomena. For example, if a heightened self-protection motivation elicited by explicit threat changes the engagement of the neural systems related to positivity bias or engages additional neural regions, this might suggest that positivity biases may represent distinct phenomena as a function of heightened self-protection motivation. On the other hand, if a self-protection motivation elicited by explicit threat does not affect the neural systems associated with positivity bias, this might suggest that positivity bias represent a unitary phenomenon regardless of whether or not self-protection motivation is especially heightened.

METHOD

Participants

Data analyses focused on eighteen right-handed participants (12 female, Age: Mean=18.7 years, SD=0.8 years) that were recruited in compliance with the human subjects regulations of the University of Texas at Austin and compensated with $15/hour or course credit. Three additional participants were excluded due to excessive head movement (+/-3mm). All participants were native English speakers, free from medications, psychological, and neurological conditions that might influence the

measurement of cerebral blood flow, and fell within a normal range of self-esteem (≥3 on a 1 to 5 scale for the Rosenberg Self-esteem Scale: Rosenberg, 1979) to avoid confounds from outlier ranges of self-esteem. As mentioned above, most individuals have positive self-views (e.g., Gray-Little, Williams, & Hancock, 1997; Twenge & Campbell, 2008); individuals with low self-esteem react to threat in an idiosyncratic manner (e.g., Vohs & Heatherton, 2004; vanDellen et al., 2011).

As described below, the present study utilized deception to provide social-evaluative feedback and studies utilizing deception methods typically find between 5-25% of participants are wise to the deception (Baumeister et al., 2005; Gardner, Picket, & Brewer, 2000; Mehta & Josephs, 2006; Stricker, Messick, & Jackson, 1969; Twenge & Campbell, 2003). Consistent with this previous research, four additional participants were excluded due to expressing suspicion in the veracity of the threat manipulation during the final debriefing procedure (two participants claimed to be suspicious prior to fMRI portion of the study, two participants became suspicious during the fMRI portion of the study; see below for Debriefing Procedure). It is unlikely that the remaining participants were suspicious but simply did not express it (Stricker, Messick, & Jackson, 1969). First, all participants were subject to a lengthy debriefing procedure (for more information, see below). Second, norms for behavioral responses have been established by previous research using the same procedure as the current study. Whereas the twenty-one participants who expressed belief in the manipulation made behavioral responses similar to those reported in previous research (Beer & Hughes, 2010; Hughes & Beer, in press-b; Vohs & Heatherton, 2004), the participants who expressed disbelief confirmed their

suspicion by responding in a manner that is inconsistent with established norms (Stricker, Messick, & Jackson, 1969).

Task

Participants completed the same social comparison task used in previous fMRI studies of social comparison (Beer & Hughes, 2010; Study 1: Hughes & Beer 2011) with the addition of a standardized social evaluative threat manipulation used in previous research (Beer, Chester, Hughes, forthcoming; Leary et al., 1998; Horton & Sedikides, 2009; Swann et al., 1990; Somerville et al., 2006, 2010a). The components of the procedure are described below followed by an overview of the task sequence.

Threat manipulation. Threat was manipulated by providing participants with unfavorable or favorable social-evaluative feedback (e.g., Beer, Chester, Hughes, forthcoming; Swann et al., 1990; Leary et al., 1998; Horton & Sedikides, 2009; Somerville et al., 2006, 2010a). Prior to scanning, participants were photographed and led to believe that other people would evaluate their likability from the photographs.

During the scanning procedure, threat was manipulated with ostensible feedback about participants' likability. Feedback Cues consisted of (a) 10 randomly-selected photographs of people (5 male, 5 female) that had ostensibly evaluated the participant's likability and (b) pie charts indicating how many people found the participant unlikable (Threat Cue condition: 6, 7, or 8 of the 10 people; No Threat condition: 0 of the 10 people)(see Figure 7).

Social comparison task. Participants rated how they compared to their average peer on personality trait words using a 5-point scale (-2=much less than the average UT student; 0=about the same as the average UT student; +2=much more than the average UT student). Trait words were undesirable traits from the narrowly-construed trait category in previous fMRI studies (e.g., stingy, jealous, messy, bossy; see Study 1A: Beer & Hughes, 2010; Study 1B: Hughes & Beer in press-a). A comparable "average peer" was ensured by recruiting University of Texas at Austin students who evaluated their personality traits in relation to the average University of Texas student of their same age and gender (Beer & Hughes, 2010; Chambers & Windschitl, 2004). Responses were reversed-scored to indicate positively-biased social comparisons (i.e., higher scores indicated the self had less of the undesirable traits compared to peers).

Task Sequence. For each trial, participants (a) received threatening or nonthreatening feedback and (b) then answered a block of social comparison questions. A Threat or No Threat Cue (4 seconds) was followed by a screen with a fixation point that indicated participants should clear their minds. The fixation point screens were randomly jittered (2 sec (50%), 4 sec (25%), 6 sec (25%); Donaldson, Petersen, Ollinger, & Buckner, 2001) to permit independent modeling of neural activation elicited by the cue and subsequent social comparison task. Participants then completed a block of 8 social-comparison questions (2 seconds each for a total of 16 seconds) followed by a screen with a fixation point that indicated they should clear their minds (16 seconds).

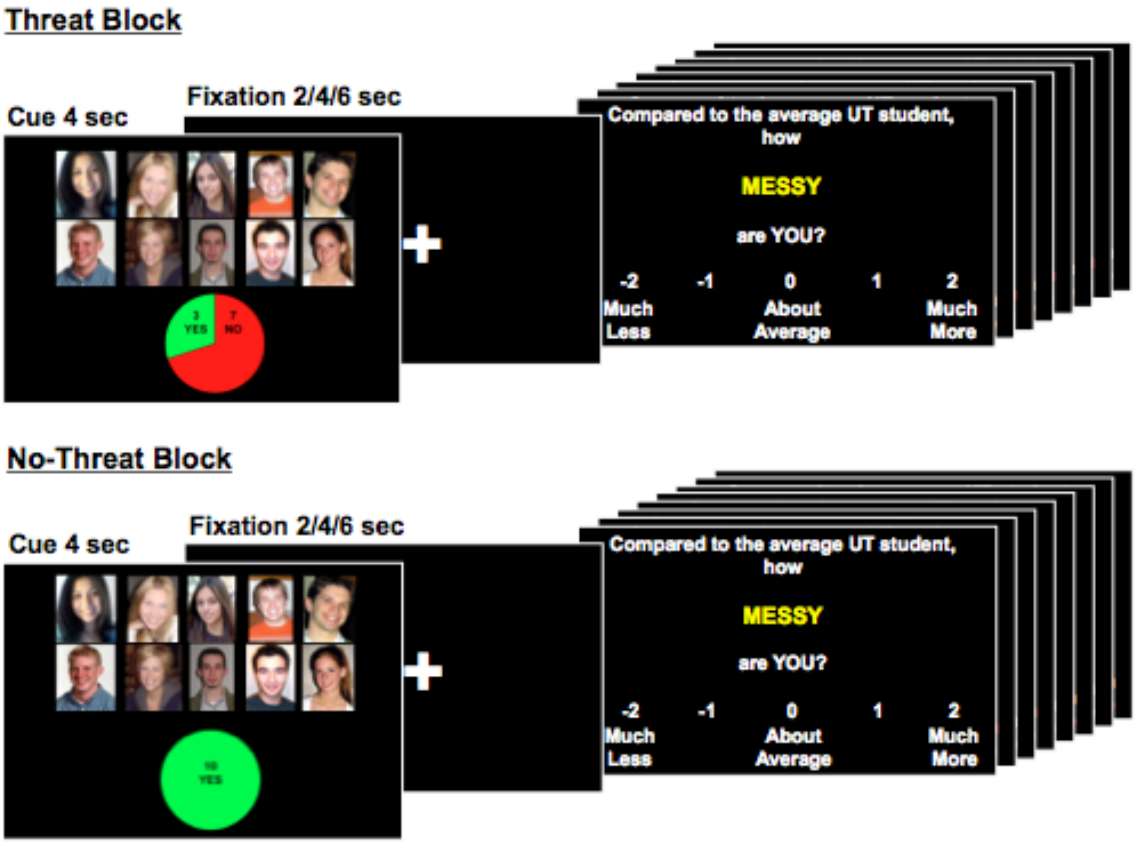

Figure 7: Stimuli and timing for threat manipulation and self-evaluation task

First, participants were presented with a 4 sec cue that indicated the Threat condition (Threat, No-Threat). Second, each cue screen was followed by a 16 sec block of the social-comparison task. Blocks were followed by a 16 sec screen depicting a fixation point. A jittered ITI separated the Threat Cue from self-evaluation Block.

FMRI data were collected in one 8-minute, 48-second scan (6 blocks of social comparison questions: 3 primed by Threat; 3 primed by No-Threat). The presentation order of the Threat and No-Threat Cues and trait words in the social comparison blocks was randomly assigned across participants. 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.

Post-Scan Task

A post-scan procedure was included to better understand how participants performed the social-comparative task. Research has robustly shown that threat's effect on social-comparative evaluations is most often accounted for by emphasizing one's own desirability rather than derogating others (Campbell, 1986; Steele et al., 1993; Aronson et al., 1995; Dodgson & Wood, 1998; vanDellen et al., 2011). There is some evidence that when people can choose the target of comparison, they may choose someone who is worse off than themselves to make themselves look better (Wills, 1981; Brown, 1986; Taylor & Lobel, 1989; Wood et al., 1999). In the present study, it was unlikely that participants could selectively derogate the target of comparison in the Threat condition. The target of comparison (i.e., the average peer of same age and gender) was held constant across the Threat manipulation. However, the possibility that the average other was evaluated differently across conditions was tested in a post-scan procedure. Participants rated the extent to which personality traits described their average peer ("How well does this trait describe the average University of Texas student:" same scale as above) in blocks that were preceded by Threat or No Threat to self.

Debriefing Procedure

In order to ensure that the participants included in all analyses were naïve to the deception involved in the threat manipulation, a thorough debriefing interview was conducted at the end of the study (for a discussion of the importance of excluding suspicious participants from analyses, see Stricker, Messick, & Jackson, 1969). This debriefing procedure was designed to give participants ample opportunity to express

whether they knew that the threatening and non-threatening feedback was false. The debriefing procedure gradually probed for participants' suspicion through open-ended questions about their general impressions of the purpose of the experiment as well as any thoughts and feelings about any aspect of the study (e.g., "What do you think the experiment was about?", "What kind of feedback did you get from other participants in the study?", "Did you have any reactions to the feedback you received?"). Participants were excluded from the final sample if they expressed suspicion or disbelief at any point during the open-ended questions of the debriefing procedure. After the question period, participants were asked not to divulge any information about any aspect of the study to their peers and provided with a full explanation of the study with a special emphasis on the bogus nature of the feedback they received (see Aronson & Carlsmith, 1968; Mills, 1976).

FMRI 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. Functional images were acquired with a GRAPPA sequence (TR=2000 ms, TE=30 ms, FOV=240, voxel size 2.5x2.5x3.3-mm) with each volume consisting of 35 axial slices oriented to the AC-PC line (e.g., Beer & Hughes, 2010; Hughes & Beer, in press-a). These parameters were implemented to optimize coverage of the orbitofrontal cortex without sacrificing whole-brain acquisition. A high-resolution SPGR T1-weighted image was also acquired from each subject.

FMRI 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. Image volumes 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 (2mm 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 period of 128-sec was applied to remove within-session drifts.

A fixed-effects analysis modeled (a) the Threat and No-Threat Social Comparison blocks using a canonical block hemodynamic response function and (b) the Threat cues and the jittered fixation screens as regressors of no interest using canonical hemodynamic response function with a temporal derivative. The 16-second fixation blocks estimated baseline for the 16-second Social Comparison blocks. A general linear model analysis created contrast images for each participant to examine neural activation in the Social Comparison blocks as a function of Threat (blocks primed by Threat vs No-Threat, No-Threat vs Threat). Contrasts from each participant were used in a second-level analysis treating participants as a random effect. The group average SPM{t} maps were masked by a priori ROIs and only clusters that survived correction for multiple comparisons ($P < 0.05$ FWE, $k = 10$) were interpreted. A priori ROIs were comprised of activation clusters found in previous studies of this social comparison task (Study 1A: Beer & Hughes,

2010; Study 1B: Hughes & Beer, in press-a): MPFC (BA 9: 12,54,34; BA 10: 8,64,24), MOFC (-2,56,-16; -4,46,-10), bilateral LOFC (left BA 47: -32,34,-14; right BA 47: 28,28,-20), dACC (10,26,34), PCC (-4,-38,28), vACC (14,38,-4), and insula (-38,14,6) (all ROIs were 8mm-radius spheres, center coordinates listed in relation to each region). The current study is the first to examine threat's effect on social comparisons so it was not possible to delineate task-specific ROIs that might relate to the threat manipulation. Instead, ROIs were delineated based on neuroanatomical regions associated with threat and its regulation in previous research (Ochsner & Gross, 2005; Kober et al., 2008): amygdala, dorsolateral PFC (DLPFC), ventrolateral PFC (VLPFC) (defined by the Automated Anatomical Labelling map: Tzourio-Mazoyer et al., 2002). Finally, correlation analyses tested whether individual differences in behavior modulated neural activation identified from the main contrasts. Parameter estimates from the main contrast were extracted (Brett et al., 2002) and correlated with individual differences in Threat's effect on positively biased social comparisons (i.e., Threat Average Social Comparison Rating minus No-Threat Average Social Comparison Rating).

RESULTS

Task Performance

In response to threat, participants significantly emphasized their own desirability. In the Threat condition, participants rated themselves as having significantly fewer undesirable traits in comparison to their average peer (Threat: M=.67, SD=.38; No-Threat: M=.55, SD=.36; $t(17)=4.07$, $p < .05$: Figure 8). Threat did not significantly affect

reaction times for social comparison ratings in the scanner (Threat: M=1.29 sec, SD= .76; No-Threat: M=1.29, SD=1.07 sec; t <1, ns) or the post-scan ratings of the average peer (Threat: M=.09, SD=.34; No-Threat: M=.18, SD=.35; t <1, ns).

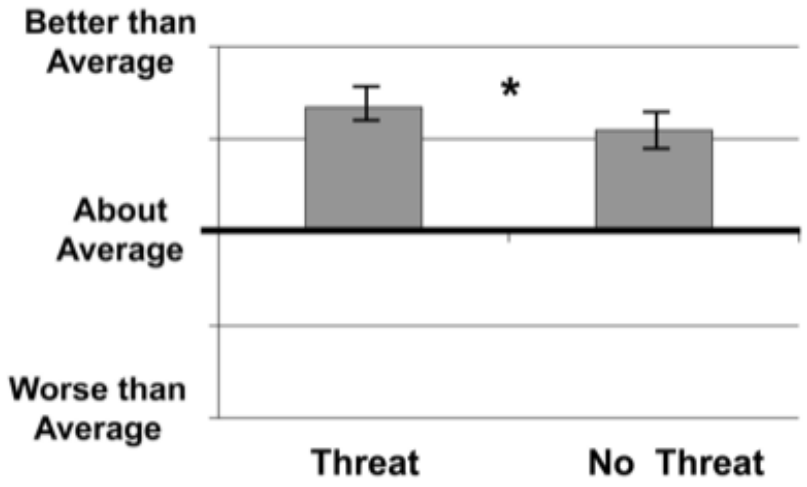

Figure 8: Behavioral results of social-comparative ratings primed by Threat and No Threat

FMRI Results

Social comparisons in response to threat are associated with *increased* activation in OFC, MPFC, amygdala, and insula

The neural activation pattern associated with positively-biased self-evaluation in the Threat condition was distinct from neural patterns found in previous neural research using this task (Study 1A: Beer & Hughes, 2010; Study 1B: Hughes & Beer, in press-a). In contrast to the No Threat condition, social comparisons in the Threat condition were associated with significantly increased activation in MOFC, bilateral LOFC, and amygdala, and, to a lesser extent, MPFC and left insula (see Table 1 and Figure 9). No

significant activation was found for the reverse contrast (Social Comparison block primed by No Threat > Threat) in any of the a priori ROIs.

Table 1: Neural regions associated with social comparisons primed by Threat versus No Threat.

			MNI Coordinates			
Brain region	BA	No. of voxels	x	y	z	t(17)
Medial OFC	11	40	-10	54	-16	6.19*
Left Lateral OFC	47	51	-30	38	-20	3.41*
Right Lateral OFC	47	158	28	32	-24	3.49*
Medial PFC	9	64	10	58	38	2.97†
Medial PFC	10	48	4	68	22	2.82†
Insula		153	-40	18	8	2.99†
Left Amygdala		90	-16	-6	-18	3.49*
Right Amygdala		133	26	2	-24	4.11*

(*) indicates $p < .05$; (†) indicates $.05 < p < .1$.

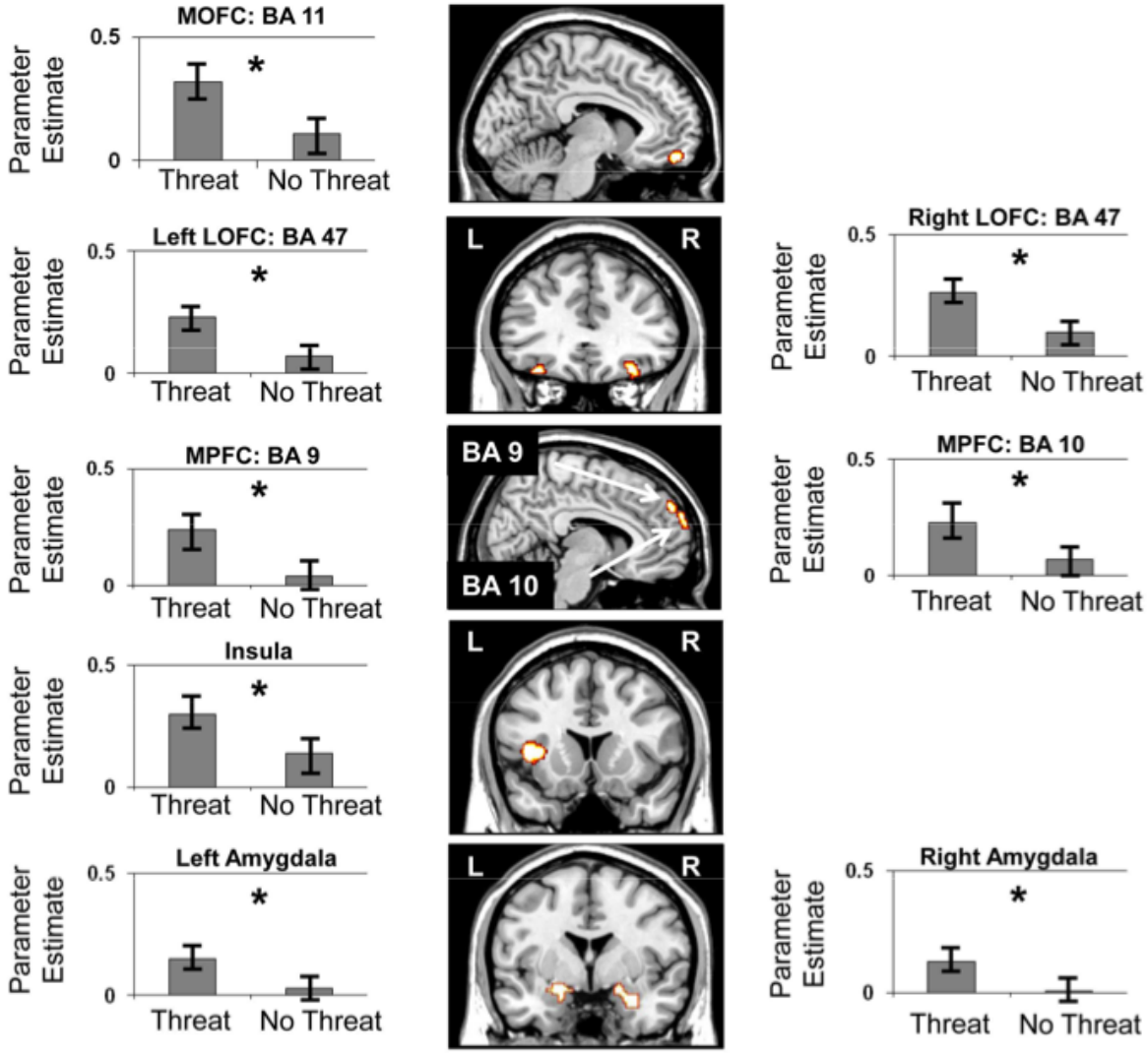

Figure 9: Neural activation from the social comparison ratings primed by the Threat versus No-Threat contrast.

Parameter estimates in relation to baseline extracted for each condition (x = -8: MOFC (BA 11); x = 34: Bilateral LOFC (BA 47); x = 6: MPFC (BA 9, BA 10); y = 16: Insula; y = 0: Bilateral Amygdala; (*) indicate significant differences.

Individual differences in positively-biased social comparisons as a function of threat modulate MOFC and MPFC

Further analyses suggested that the increased MOFC and MPFC activation in the contrast of Social Comparison block primed by Threat > No Threat was driven by individual differences in positively-biased social comparisons. The more participants evaluated themselves as 'above average' in the Threat condition, the more they recruited the MOFC and MPFC regions identified in the main contrast. Individual differences in positively-biased social comparisons as a function of Threat significantly correlated with parameter estimates extracted from the MOFC ($r = .78$, $p < .05$) and MPFC ($r = .53$, $p < .05$) activity found in the Social Comparison block primed by Threat > No Threat contrast (see Figure 10).

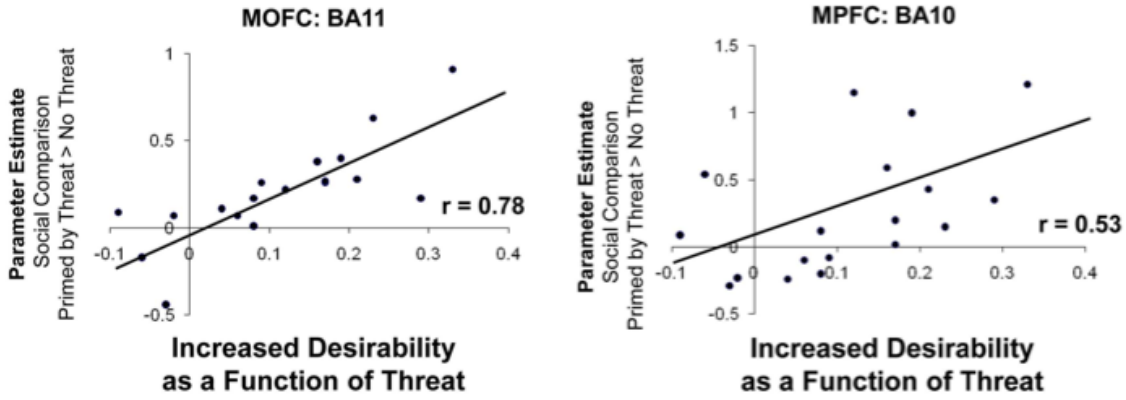

Figure 10: Individual differences in positively-biased evaluations as a function of Threat modulate neural activation

Social-comparative ratings modulate MOFC and MPFC activation from the Threat > No-Threat contrast. Scatterplots depict individual differences in positively-biased social comparisons as a function of Threat (Social Comparison ratings primed by Threat – Social Comparison ratings primed by No-Threat) in relation to parameter estimates of the

MOFC and MPFC activation identified in the Social Comparison primed by Threat > Social Comparison primed by No-Threat contrast.

DISCUSSION

Study 2 sought to extend the findings from Studies 1A and 1B by examining whether a heightened self-protection motivation changes the engagement of a core set of neural regions implicated in positively-biased self-evaluation. In Study 2, an explicit threat manipulation was used to elicit positivity bias in the same social comparison task used in Studies 1A and 1B. Study 2 shows for the first time that an explicit threat manipulation changes the engagement of neural systems associated with positivity bias, and engages additional neural regions. Specifically, social comparison ratings in response to threatening feedback were associated with *increased* MOFC, LOFC, amygdala, and, to a lesser extent, MPFC and insula activation. Furthermore, individual differences in the extent to which threat increased positively-biased self-evaluations were associated with increased MOFC and MPFC activation. In contrast, Study 1A and 1B found that positively-biased evaluations elicited by contextual factors other than an explicit threat were associated with *decreased* activation in MOFC, LOFC, and, to a lesser extent, dACC. The similarities and differences between the neural associations of positivity bias in the current study and previous research cannot be accounted for by differences in the stimuli used to measure positivity bias; the current study used the same social comparison task and stimuli as Study 1A and Study 1B. The major difference was that the current study elicited positivity bias as a response to an explicit manipulation of threat.

While Study 2 represents a step toward informing different perspectives for how positivity biases in self-evaluations are accomplished, the pattern of neural activation associated with positivity bias when threat is explicitly heightened does not conclusively address whether positivity biases represent a unitary phenomenon or multiple phenomena as a function of heightened self-protection motivation. Study 2 did not find the same reduction in OFC and dACC activation associated with positivity bias when threat was not explicitly heightened (Study 1A and 1B). Study 2 also did not find that a completely different set of neural regions were involved in eliciting positivity bias when threat was explicitly heightened. Instead, Study 2 found a different pattern of neural activation in many of the same regions as Studies 1A and 1B (i.e., MOFC, LOFC), as well as additional neural modulation (i.e., MPFC, amygdala, insula). Taken together, these findings raise questions that promise to further elucidate how positivity biases are underpinned in the brain, which may in turn lead to a better understanding of how they are accomplished in different motivational contexts.

First, what role does MOFC play in positivity bias that might help explain why the direction of its association changes as a function of a self-protection motivation? One possibility may be that the MOFC supports the same psychological process regardless of whether threat is explicitly heightened, but may produce attenuation versus enhancement of positively-biased responses in different contexts. In this case, MOFC may support a common psychological process in positivity bias but the implementation of that psychological process may differ depending on contextual motivators. An alternative possibility is that different psychological processes are supported by the MOFC in

positively-biased evaluation. For example, MOFC may interact with additional neural regions that are engaged by a heightened self-protection motivation, and its unique interaction may support a psychological process that is distinct from the psychological process supported by MOFC when self-protection motivation is not heightened. In this case, MOFC's interaction with one network of regions may engage a mechanism that produces enhancement of positively-biased responses in the context of explicit threat, and its interaction with a second network of regions may engage a different mechanism that produces attenuation of positively-biased responses when threat is not explicitly heightened. Research that addresses these possibilities is necessary in order to begin to shed light on whether self-protection motivation engages a unique path to positivity biases in self-evaluation.

Second, what is the role of MPFC in positivity bias in the context of threat? One possibility is that the correlation between MPFC and MOFC activation and increased positivity bias after threat may indicate that MPFC and MOFC interact and their joint activation leads to increased positivity bias in evaluations when threat is explicitly heightened. A related possibility is that MPFC engages a mechanism that operates in parallel or in addition to the mechanism supported by MOFC, and their parallel or additive activation leads to increased positivity bias when threat is explicitly heightened. For example, people compensate for threat by drawing on core aspects of self and increasing their influence in subsequent self-evaluations (Aronson et al., 1995; Baumeister & Jones, 1978; Dodgson & Wood, 1998; Kunda, 1990; Steele et al., 1993; vanDellen et al., 2011; Wood et al., 1999), and there are reasons to believe that the

MPFC may support access to these core aspects of self. Previous research suggests that MPFC is associated with accessing self-related information in general (e.g., Macrae et al., 2004; Moran et al., 2005; Ochsner et al., 2005), and more certainly held self-related information in particular (D'Argembeau et al., 2012). However, more research is needed to examine the precise mechanism that is supported by the MPFC in positively-biased evaluation in the context of threat.

Lastly, more research is needed to understand the role of the additional neural regions that are brought online in the context of explicit threat. For example, amygdala and insula activation were greater for social comparisons in response to explicit threat, but their activation did not predict individual differences in positivity bias as a function of explicit threat. However, this does not discount the possibility that amygdala and insula activation may play a role in positively-biased evaluation as a function of explicit threat. One possibility is that these additional neural regions may modulate activation in the MOFC and MPFC network that predicts individual variability in positively-biased responses as a function of threat. Another possibility is that cross-subject correlations aren't telling the whole story about amygdala and insula function. Although amygdala and insula were not linked to positively-biased responses across subjects, they may be linked to positively-biased responses within subjects. However, the better-than-average paradigm and the blocked-nature of the task used are not well suited to examine the relationship between trial-by-trial neural activation and positively-biased responses. Future research will benefit from implementing tasks that allow a trial-by-trial mapping

of neural activation and positively-biased responses as a function of explicit threat to uncover potential relationships at a within subject level of analysis.

Finally, a limitation of Study 2 is that the within subjects nature of the threat manipulation did not allow for an increased number of blocks, which may have lead to decreased statistical power. Specifically, pilot testing revealed that the inclusion of a greater number of blocks reduces the effect of threat on behavioral responses, potentially due to habituation to the threat cues. While the blocked nature of the task may have partially alleviated power issues related to the limitation of the threat manipulation, it does not discount the possibility that a heightened self-protection motivation may engage additional neural regions that we did not have the power to detect. Therefore, future research should implement task designs that allow for increased repetition of the threat manipulation in order to increase statistical power.

AIM 3: WHAT MECHANISMS ARE SUPPORTED BY NEURAL SYSTEMS ASSOCIATED WITH POSITIVITY BIAS?

Study 3

INTRODUCTION

Study 3 examines one possibility raised by Studies 1A, 1B and Study 2 that may explain why MOFC supports increased positivity bias in self-evaluations when threat is explicitly heightened and decreased positivity bias in self-evaluations when threat is not explicitly heightened: shifts in decision thresholds. Researchers have theorized that one way to change the positively-biased nature of evaluations is by shifting the decision thresholds that influence the expression of baseline positively-biased associations (Paulhus et al., 2003). Research suggests that positivity bias is prepotent or at least relies on relatively automatized processing (Alicke et al., 1995; Beer & Hughes, 2010; Beer, Chester, & Hughes, forthcoming; Hixon & Swann, 1993; Hughes & Beer, unpublished data; Koole et al., 2001; Lench & Ditto, 2008; Paulhus et al., 1989; Swann et al., 1990). Most people have positive associations with self (Gray-Little et al., 1997; Koole et al., 2001; Twenge & Campbell, 2008). Moreover, people's positively-biased tendency to claim that their personality characteristics are more desirable than the personality characteristics of their average peer is exacerbated (e.g., Beer & Hughes, 2010; Hixon & Swann, 1993; Lench & Ditto, 2008; Paulhus et al., 1989; Swann et al., 1990) or unaffected by cognitive load (e.g., Alicke et al., 1995). Therefore, one possibility is that the association between MOFC activation and positively-biased responses might reflect a departure from this baseline or automatic positively-biased tendency as contexts change.

In Study 1A and 1B, people's positively-biased tendency to claim that they or their romantic partners have more desirable personalities than their peers was attenuated when the comparisons were made for narrowly-construed traits; attenuation as a function of trait breadth was predicted by MOFC activation. The increased MOFC activation may reflect a shift in decision threshold because one potential effect of a trait breadth manipulation is that it shifts how liberally one can construe a trait as self-relevant and, consequently, judge oneself as better than one's peers on that trait (Dunning et al., 1989). Therefore, the increased MOFC activation may reflect the extent to which participants were sensitive to the more conservative decision threshold inherent in claiming narrowly-construed traits as especially self-relevant. In Study 2, people tended to respond to explicit threat by increasing their baseline self-serving tendency to claim that they have more desirable personalities than their peers as a way to compensate for threat; increases as a function of threat were predicted by MOFC activation. The increased MOFC activation may reflect a shift towards more liberal decision thresholds for expressing positive information about the self in the face of threat, and subsequently, increased positively-biased responses.

The suggestion that MOFC activation may support shifts in decision thresholds as contexts change is consistent with the more general role of the OFC in contextual updating. Research has demonstrated a broad role of MOFC in contextual updating, that is, in shifting the threshold at which prepotent tendencies are expressed as contexts change (Beer, Shimamura, & Knight, 2004; Lhermitte, 1986; Stuss & Benson, 1984). Medial OFC activation has been associated with both the downregulation and

upregulation of behavior in relation to contextual changes (e.g., Bhanji & Beer, in press; Cooney et al., 2011; Hare, Malmaud, & Rangel, 2011; Mehta & Beer, 2009). Conversely, damage to the OFC, including the medial OFC, is classically associated with the inflexible expression of baseline or automatic tendencies even when they become contextually inappropriate (Lhermitte, 1986; Stuss & Benson, 1984 and see Beer et al., 2006; Fellows & Farah, 2003; Sellitto, Ciaramelli, & di Pellegrino, 2011). Therefore, one way MOFC may affect self-judgments is by updating their underlying baseline (or automatic) strategies or components.

Study 3 tests the possibility that MOFC supports changes in decision thresholds in positively-biased evaluation by drawing on signal detection measurement of decision thresholds and an accountability manipulation known to shift the relative liberality of decision thresholds that underlie positively-biased responses (Paulhus et al., 2003). Accountability refers to the implicit or explicit 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 & Tetlock, 1999; Sedikides et al., 2002). 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 & Swets, 1966; Macmillan & Creelman, 1991). Previous research on the effect of accountability on overclaiming has applied signal detection theory (SDT) techniques to claims of knowledge and found that warning participants that information may be nonexistent reduces inflated claims of knowledge by

shifting decision thresholds in a more conservative direction (Paulhus et al., 2003). 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 positively-biased responses and test its relation to OFC activation. OFC activation should predict the extent to which participants adopt more conservative (i.e., less positively-biased) decision thresholds when inflated claims have the potential to be exposed. Results of this study are also reported in a published manuscript (Hughes & Beer, in press-b).

METHOD

Participants

Eighteen right-handed participants (9 female, 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/hour 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.

Task

Participants completed a modified version of the over-claiming questionnaire and accountability manipulation used in previous research (Paulhus et al., 2003). Participants rated their familiarity with blocks of knowledge items in two Accountability conditions (Accountable, Unaccountable: see Figure 11). Blocks of items were preceded by task

instructions (4 secs) 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). 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 secs 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.

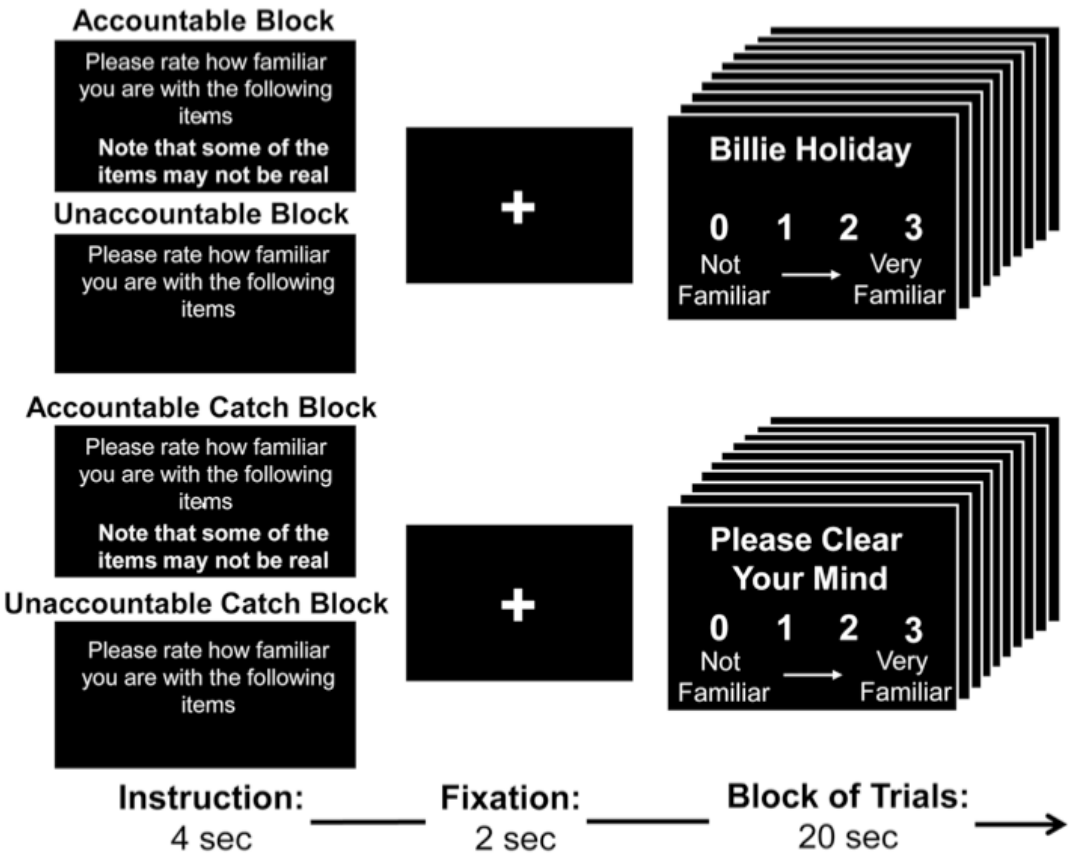

Figure 11: Stimuli and timing in over-claiming bias task

Participants saw an instruction screen that either did or did not provide an Accountability cue (i.e., some of the items may be nonexistent), 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.

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 versus

Unaccountable blocks were not merely due to the presence of the Accountability cue screen (see Figure 11). 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 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 positively-biased 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 & Swets, 1966) was used to model two indices: shifts in participants' tendency to make inflated 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 positivity bias 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 Figure 12A). 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 inflated claims (see Figure 12B). Discriminability (*d'*) does not necessarily reflect the positively-biased 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 inflated 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 Figure 12A).

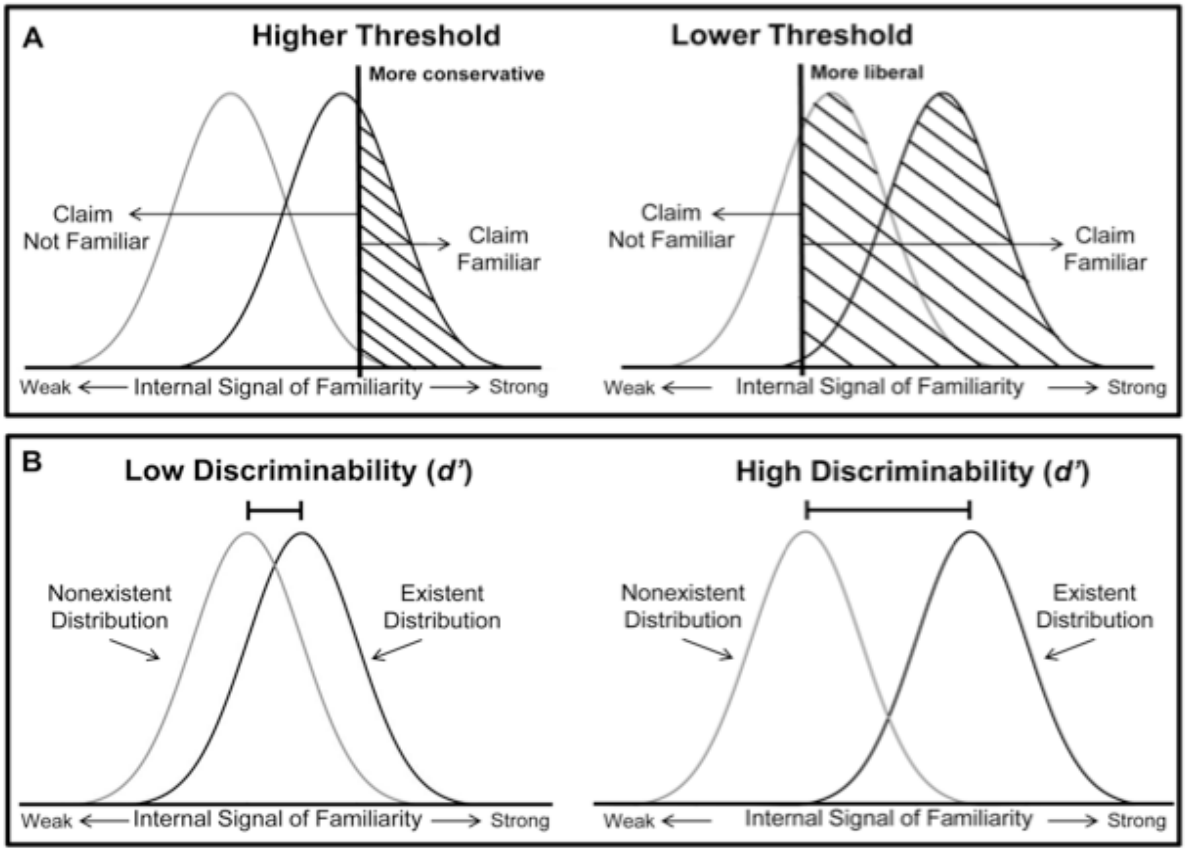

Figure 12: Examples of familiarity distributions and decision thresholds for existent and nonexistent items

(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).

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 existent 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 positively-biased nature of knowledge claims because it reflects the strength of the evidence needed in order to claim 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 (see Figure 12A). Therefore, the decision threshold (*c*) was used to measure over-claiming and its shift measured by contrasting across accountability conditions (Macmillan and Creelman, 1991; Paulhus et al., 2003):

$$criterion = -\frac{z(HR) + z(FAR)}{2} \quad (1)$$

On the other hand, SDT takes the perspective that discriminability (*d'*) reflects the ability to discriminate between existent and nonexistent items, rather than the positively-biased 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 (see Figure 12B). 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) on average are theorized to generate a stronger sense of familiarity than nonexistent items (e.g., J.D. Louis). However, SDT also suggests that an observer may sometimes not feel familiar with an existent item 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 Figure 12). 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):

$$d' = z(HR) - z(FAR) \tag{2}$$

FMRI 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. 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. A high resolution SPGR T1-weighted image was also acquired from each subject.

FMRI 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. Image volumes 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 seconds 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 inflated claims of knowledge after being held Accountable because participants rely less liberally on easily accessible familiarity cues (i.e, 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 inflated claims of knowledge after receiving an Accountability cue (see Table 2). 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* inflated) after being cued that inflated 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= 0.38, SD= 0.14, $t(17)= 3.81$, $p < .05$) and Unaccountable (M= 0.48, SD= 0.16, $t(17)= 5.81$, $p < .05$) blocks. Lastly, Participants' raw familiarity ratings on average were significantly higher for real items compared fake items in both the Accountable (real items: M = 1.13, SD = .40; fake items: M = .23, SD = .19, $t(17)= 8.65$, $p < .05$) and Unaccountable (real items: M = 1.45, SD = .48; fake 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 sec, SD = .12 sec; Unaccountable RT: M = 1.18 sec, SD =.10 sec; $t(17) = 1.34$, ns).

Table 2: Task performance in the Accountable and Unaccountable blocks.

	Accountable		Unaccountable		
	M	SD	M	SD	t(17)
Decision threshold (*c*)	.91	.25	.79	.28	2.46*
Discriminability (*d'*)	1.07	.52	1.44	.54	2.60*
Hit rate (average)	.38	.14	.48	.16	-4.49*
False alarm rate (average)	.08	.05	.08	.05	.30
Reaction time (seconds)	1.22	.12	1.18	.10	1.34

(*) indicates $p < .05$.

FMRI Results

Conditions that reduce over-claiming responses are associated with increased OFC, MPFC, and dACC activation

Consistent with previous research on positively-biased evaluation (Blackwood et al., 2003; Krusemark et al., 2008; Beer et al., 2010; Beer & Hughes, 2010; Hughes & Beer, in press-a), the current study on inflated 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 3, Figure 13). 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 4 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.

Table 3: Neural regions associated with judgments in the Accountable versus Unaccountable contrast.

		MNI Coordinates				
Region of Activation	BA	x	y	z	T-stat	Cluster size
Medial orbitofrontal (L)	11	-6	58	-20	6.62	226
		-12	48	-18	6.31	
		-4	44	-24	5.05	
Lateral orbitofrontal (R)	47	32	44	-18	5.75	208
		38	56	-16	4.59	
Lateral orbitofrontal (L)	47	-30	56	-12	6.96	717
		-34	40	-18	6.88	
Medial prefrontal (L)	10	-16	64	-2	6.12	262
		-10	70	10	5.58	
		0	68	18	4.88	
Medial prefrontal (L)	8/9	-2	34	62	4.96	235
		0	56	44	4.80	
		0	60	32	3.96	
Dorsal anterior cingulate cortex (L)	24	-4	14	34	5.44	174

0	6	26	4.06	
0	-8	30	3.40	

BA = Brodmann's Area; all regions are significant at P < .05 corrected for multiple comparisons at the cluster-level.

Table 4: Neural regions associated with Accountable "Catch" Blocks versus Unaccountable "Catch" Blocks.

		MNI Coordinates				
Region of Activation	BA	x	y	z	T-stat	Cluster size
Occipital (R)	18/19	26	-96	14	8.98	1294
		32	-80	-6	6.02	
		24	-84	-6	5.81	
Occipital (L)	18	-18	-92	-8	7.09	1324
		-28	-72	-10	6.83	
		-20	-78	-8	6.77	
Lateral prefrontal (R)	45	50	28	2	6.27	515
		48	20	0	5.56	
		58	24	10	5.41	
Inferior parietal (L)	40	-58	-42	50	5.56	183
		-56	-50	42	3.45	

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.

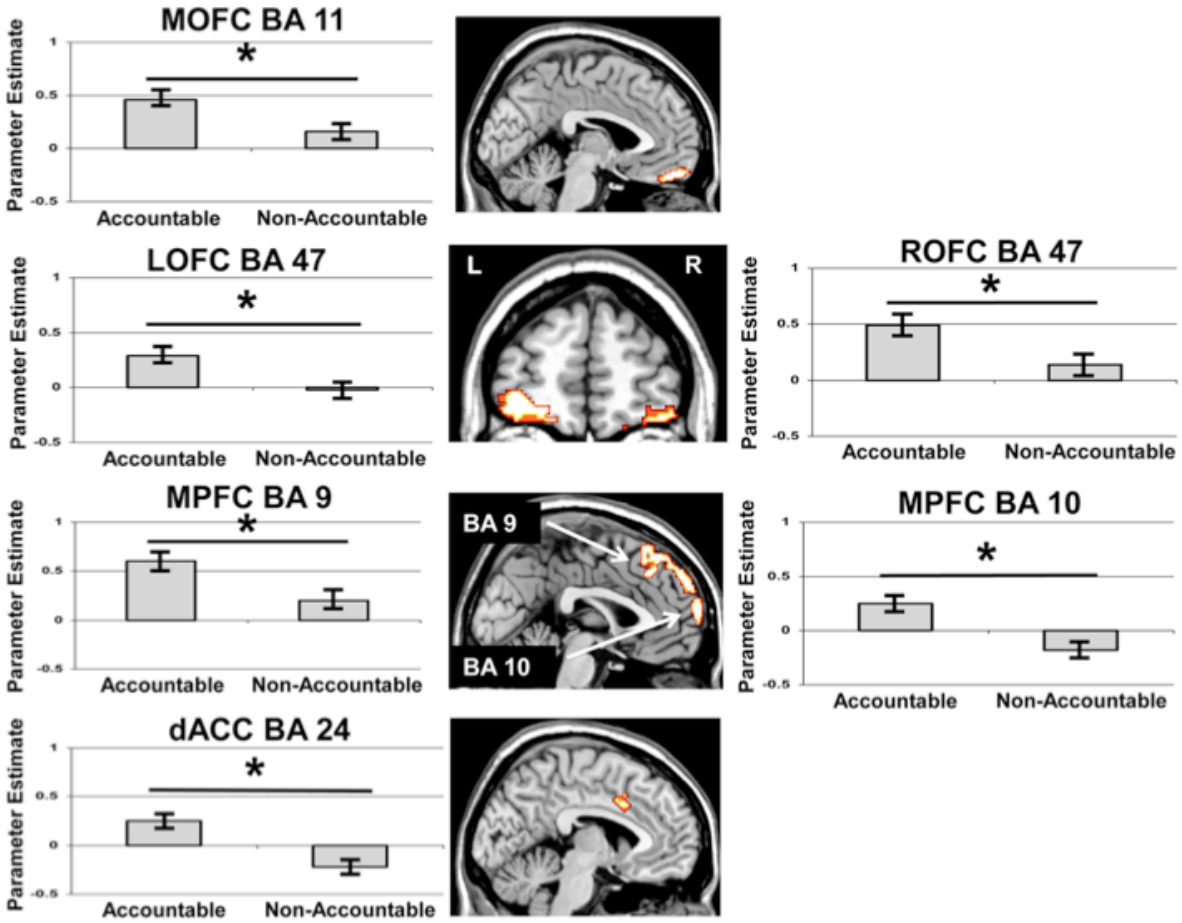

Figure 13: Neural activation defined by the Accountable versus Unaccountable contrast.

Parameter estimates plotted in relation to baseline for each condition. (*) indicates $p < .05$. MOFC = medial orbitofrontal cortex; LOFC = left lateral orbitofrontal cortex; OFC = right lateral orbitofrontal cortex; MPFC = medial prefrontal cortex; dACC = dorsal anterior cingulate cortex.

Individual differences in shifts in decision threshold modulate MOFC

Of 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 Figure 14). Specifically,

activation in the MOFC region identified in the Accountable > Unaccountable contrast predicted the extent to which participants adopted a more conservative decision threshold as a function of Accountability (r= 0.52, p <.05, controlling for discriminability; see Figure 14). In other words, MOFC activation counteracted over-claiming responses by significantly predicting the extent to which participants shifted their decision threshold in a conservative manner across contexts.

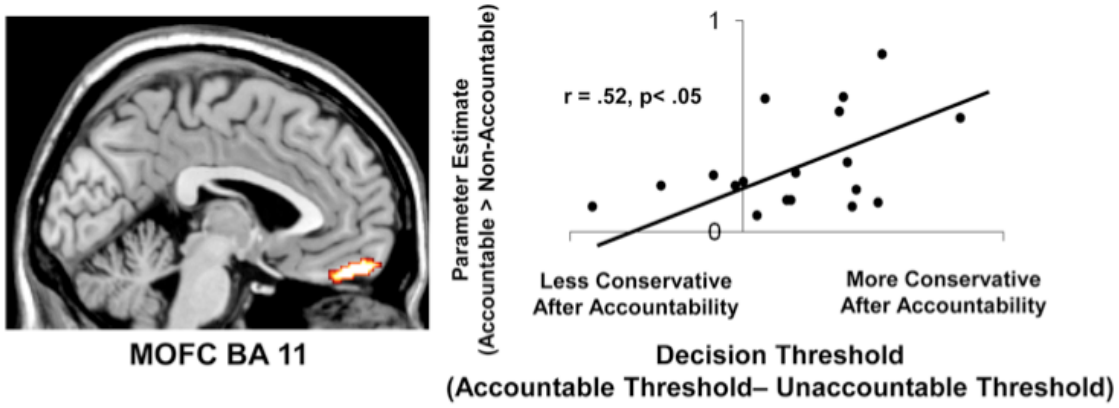

Figure 14: Individual differences in conservative decision threshold shifts (*c*) modulate MOFC activation.

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.

DISCUSSION

Study 3 addresses the possibility that MOFC may support a common psychological process in positively-biased evaluation regardless of whether threat is explicitly heightened. The previous studies found robust evidence for an association

between MOFC function and positivity bias, but very little is understood about the psychological processes that account for this association. Study 3 provided a direct test of the hypothesis that MOFC may support a common psychological process in positivity bias by drawing on established behavioral methods that combine signal detection measurement of decision thresholds and a manipulation known to shift decision thresholds underlying positively-biased responses. Study 3 found evidence that MOFC activation supports shifts in decision thresholds that influence the expression of positivity bias in evaluations. Specifically, participants recruited more MOFC activation to the extent that they shifted their decision threshold in a more conservative (i.e., less positively-biased manner) when held accountable. This finding may explain why MOFC activation supported reduced positivity bias when threat was not explicitly heightened (Studies 1A and 1B) and increased positivity bias when threat was explicitly manipulated (Study 2). It may be that MOFC supports a common decision threshold shift function in positively-biased evaluation, but people may engage this psychological process differently depending on whether a self-protection motivation is especially heightened.

While the findings from Study 3 are consistent with the possibility that MOFC supports a common mechanism in positively-biased evaluation, more research is needed to determine whether MOFC predicts both liberal and conservative shifts in decision thresholds in different contexts. In the current study, 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 automatic tendency to construe weak familiarity cues as indicative of actual knowledge. 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. However, if the association between MOFC engagement and positively-biased 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. This possibility is consistent with the findings from Study 2, in which MOFC activation predicted increases in positively-biased self-evaluations. The increased MOFC activation may have reflected a shift towards more liberal decision thresholds for expressing positive information about the self as a way to cope with threat. However, more research is needed to directly test whether MOFC activation in positively-biased evaluation also predicts liberal shifts in the decision thresholds for expressing baseline positively-biased associations. 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 positively-biased responses (Sedikides et al., 2002; Paulhus et al., 2003). One possibility is to examine decision threshold shifts in an overclaiming paradigm in which participants are instructed to make a good impression on others by appearing especially intelligent (e.g., Paulhus et al., 2003) or by providing participants with threatening feedback about their intelligence (e.g., Schmeichel &

Demaree, 2010). In these cases, MOFC activation may predict more liberal shifts in decision thresholds and increased positivity bias.

Another important direction for understanding the neural underpinnings of positivity bias is to understand why lateral OFC, dACC and MPFC tend to show differential activation in relation to conditions of positively-biased evaluation but do not always significantly predict individual differences in positively-biased evaluation (see Beer & Hughes, 2010; Hughes & Beer, in press-a; Krusemark et al., 2008). Research in other domains has associated many of these neural regions with difficulty or conflict (e.g., Beer, Shimamura, & Knight, 2004; Botvinick, Cohen & Carter, 2004; Grinband et al., 2011) yet that account does not readily fit with the existing research on positivity bias. For example, if conditions that exacerbate or attenuate positivity bias 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., Hughes & Beer, in press-a; Hughes & Beer, forthcoming). 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, dACC, and MPFC 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 positivity bias.

Finally, a limitation of Study 3 is that the within subjects nature of the accountability manipulation may have not allowed for an increased number of blocks,

which may have lead to decreased statistical power. While the blocked nature of the task may have partially alleviated power issues related to the within subjects accountability manipulation, it does not discount the possibility that lateral OFC, dACC, and MPFC may have also predicted individual differences in decision threshold shifts or discriminability. In addition, 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 present work 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).

GENERAL DISCUSSION

Overview of Findings

The work presented in this dissertation represents a first step toward examining the neural mechanisms that underlie positivity biases in order to inform our understanding of how positivity biases are accomplished as a function of different motivational contexts. Previous behavioral research is limited in its ability to resolve how positivity bias occurs in different contexts because behavioral indices of positivity bias are similar in both threatening and non-threatening contexts. We used an interdisciplinary approach that combined established experimental methods and theory from social psychology, decision-making, and neuroscience in an attempt to peer underneath the hood of positivity bias in different motivational contexts. Specifically, the four experiments presented here aimed to (1) identify a core network of neural regions that is associated with positivity bias (Studies 1A and 1B), (2) examine how a heightened self-protection motivation elicited by threat changes the engagement of that core network of neural regions (Study 2), and (3) begin to test the specific mechanisms that may be instantiated in those regions (Study 3). In doing so, this body of work begins to establish a strong empirical foundation for the neural systems of positivity biases in evaluations and sets the stage for future work aimed at uncovering the specific mechanisms instantiated in those systems as well as how those systems may interact.

The first pair of experiments sought to identify a core set of neural regions that underlie positivity biases (Studies 1A and 1B). In particular, these studies used functional

magnetic resonance imaging to examine the neural systems involved in a commonly used indicator of positivity bias, namely, the better-than-average effect. We found evidence for a core set of neural regions comprised of the OFC and, to a lesser extent dACC that was negatively modulated by better-than-average judgments. These findings contribute to a large body of neural research on self-evaluation and social cognition by identifying for the first time a core network of neural regions involved in positivity biases in self-evaluation and social cognition.

Study 2 extended this line of research by explicitly heightening a self-protection motivation and examining its effect on the core network of neural regions that underlie positively-biased evaluation. Interestingly, Study 2 demonstrated that explicit threat changed the engagement of the neural systems associated with positivity bias as well as engaged additional neural modulation. In particular, positively-biased evaluations in response to threatening feedback were associated with *increased* activation in a common set of regions comprised of MOFC, LOFC, and MPFC, with the additional recruitment of increased amygdala and insula activation. Furthermore, MOFC and MPFC activation predicted individual differences in increased positivity bias as a function of threat. The different pattern of activation in an overlapping set of regions and additional neural modulation as a function of threat represents a first step toward understanding whether positively-biased evaluations engage distinct psychological processes as a function of self-protection. However, more research is needed to better understand whether the neural findings suggest the presence of common or distinct mechanisms in positively-biased evaluation as a function of a heightened self-protection motivation. In particular,

the MOFC appears to be a critical region in positively-biased judgment that shows opposite patterns of activation as a function of threat across studies.

Study 3 tested the specific hypothesis that MOFC supports a common mechanism in positively-biased judgment by drawing on signal detection techniques to model decision threshold shifts that may underlie positively-biased responses. Study 3 found evidence that MOFC activation supports shifts in decision thresholds that influence the expression of positively-biased responses as a function of the motivational context. This finding suggests that MOFC may be a convergence zone for positively-biased judgment in the brain and helps explain why MOFC activation leads to reduced positivity bias when self-protection is not especially heightened and increased positivity bias when self-protection is heightened. Specifically, MOFC may support a common mechanism in positivity bias, but the common mechanism may be implemented differently as a function of the motivational context. For example, people may implement the common mechanism differently as a way to increase positively-biased evaluations in the context of explicit threat. If people implement a common mechanism in different ways as a function of the motivational context (e.g., self-protection), then this finding may provide some preliminary evidence that positivity biases may reflect distinct phenomena as a function of self-protection motivation.

Taken together, these studies represent a first step toward developing neural models of positively-biased judgment in order move away from underspecified behavioral comparisons and begin to understand how positivity biases are accomplished in different motivational contexts. Specifically, the studies provide evidence that (1)

MOFC may be a convergence zone for positively-bias judgment that supports a common decision process that is implemented differently as a function of the motivational context, (2) MPFC may play a distinct role in compensating for threat by increasing positivity bias, and (3) Amygdala and Insula may be involved in positivity bias in the context of threat, but their role remains underdetermined. The role of each of these neural regions and their implications for whether positivity bias reflects multiple processes depending on contextual motivators can be informed by a consideration of their neuroanatomy and the functions they may support.

Role of Medial Orbitofrontal Cortex in Positivity Bias

NEUROANATOMY OF THE MOFC

The neuroanatomy of MOFC is consistent with the present proposal that the MOFC may support a decision process in positivity bias that is sensitive to motivational contexts. The MOFC is relatively unique among regions of the cortex because it is densely interconnected with structures involved in affective processing, higher-level cognition, and autonomic regulation (for review see Price & Drevets, 2010). Evidence for affective input comes from MOFC's interconnection with the amygdala and ventral striatum, two structures important for emotional and reinforcement learning, and from MOFC's interconnection with the thalamus, a structure that receives excitatory and inhibitory inputs from amygdala and basal ganglia (Amaral & Price, 1984; Cavada et al., 2000; Price & Drevets, 2010). These cortico-striato-pallido-thalamic loops are important

for maintaining a course of action and for switching to a new course of action when previously advantageous behaviors become disadvantageous (Price & Drevets, 2010). Evidence for higher-level cognitive inputs come from MOFC's interconnection with other cortical areas, including the lateral OFC (LOFC BA 47) region involved in processing hedonic aspects of sensory information (Kringelbach & Rolls, 2004; Schoenbaum et al., 2011), dorsolateral prefrontal regions involved in goal formation and maintenance (Goldman-Rakic, 1987; Fuster, 2001), and MPFC regions important for attending to internally generated information (BA 10 and 9, see MPFC section below). Lastly, evidence for autonomic control comes from MOFC's outputs to the hypothalamus, periaqueductal gray (PAG), and other regions important for regulating autonomic activity and preparing the body for action (Ongur & Price, 2000; Price & Drevets, 2010). The pattern of connectivity suggests that MOFC may be uniquely positioned to integrate basic affective, reinforcing, and interoceptive signals, higher-level cognitive representations, and goals in order to flexibly guide decision-making (Kringelbach & Rolls, 2004; Roy, Shohamy, & Wager, 2012; Wallis, 2007).

FUNCTIONS OF THE MOFC

As suggested by its neuroanatomy, MOFC function is not limited to positively-biased judgment, but rather plays a more general role in flexible and goal-directed decision-making (Roy, Shohamy, & Wager, 2012). Some of the earliest hints of the role of MOFC in decision-making come from clinical case studies showing that patients with OFC damage exhibit fundamental impairments in decision-making (Lhermitte, 1986;

Rolls et al., 1994; Stuss & Benson, 1984). In particular, damage to the OFC, including the MOFC, is classically associated with the inflexible expression of prepotent or automatic tendencies even when they become contextually inappropriate (Bechara et al., 2000; Beer et al., 2006; Fellows & Farah, 2003; Lhermitte, 1986; Rolls et al., 1994; Stuss & Benson, 1984; Sellitto et al., 2011). Patients with OFC damage are able to state the rules that govern a task, but then fail to apply those rules to their behavior. For example, even though patients are aware of the different rules that govern interactions with strangers versus close friends, patients still talk to strangers as though they are close friends (Beer et al., 2006).

The classic observation that patients with OFC damage are impaired in their ability to flexibly adjust behavior raises the possibility that MOFC may play a role in certain aspects of reinforcement learning. Flexible decision-making requires the ability to estimate the values of actions based on previous experience and subsequently decide between them. Reinforcement learning models are used to examine how this process is underpinned in the brain (Sutton & Barto, 1998). In this formalism, when an action yields more or less reward than expected, the action-value estimate is updated to guide future decisions. MOFC lesions in humans and other species have been shown to impair the ability to update the decision-making policy when contingencies change and previously successful strategies are no longer advantageous (e.g., reversal learning and fear extinction: Balleine & O'Doherty, 2010; Bechara et al., 2000; Fellows & Farah, 2003; Izquierdo et al., 2004; Milad & Quirk, 2002; Jones & Mishkin, 1972). Similarly, a number of fMRI studies in healthy individuals find evidence that MOFC activation is

involved in tracking changes in the reward contingency between actions and outcomes, which is necessary to flexibly update behavior (Behrens et al., 2008; Daw et al., 2006; Glascher et al., 2009; Hampton et al., 2006; Tanaka et al., 2008; Valentin et al., 2007). In addition, the burgeoning literature on "model-based" reinforcement learning – a more sophisticated form of choice that utilizes the structure of the environment – suggests that the MOFC may be involved in evaluating and updating internal representations of the structure of the structure of the environment in order to flexibly guide decisions (Daw et al., 2011; Hampton et al., 2006; Summerfield & Koechlin, 2008; Takahashi et al., 2011). However, it is not known if the decision processes supported by MOFC in reinforcement learning are sensitive to changes in motivational contexts as in the present work, since previous research has not examined reinforcement learning processes as a function of different contextual motivators.

While the reinforcement learning literature has not manipulated motivational contexts, research on goal-directed decision-making shows that MOFC is sensitive to changes in motivational contexts. For example, there is evidence that the association between MOFC activation and the value of a stimulus is modulated by subjective preferences (e.g., Pepsi vs. Coke) and contextual factors (e.g., wines labeled as expensive vs. inexpensive)(Arana et al., 2003; de Araujo et al., 2005; Fellows & Farah 2007; McCabe et al., 2008; McClure et al., 2004; Plassmann et al., 2008), changes in the motivational significance of a stimulus (e.g., chocolate) as a function of satiety (Gottfried et al., 2003; Kringelbach et al., 2003; Small et al., 2001; Valentin et al., 2007), and

changes in the motivational significance of options after integrating higher-order versus prepotent goals (e.g., be healthy vs. eat tasty: Bhanji & Beer, in press; Hare et al., 2011).

The present findings extend research on the decision processes supported by MOFC by conceptualizing a role for this region in biased social judgment as a function of contextual motivators. One interpretation of the present findings is that MOFC may support a common decision process in positively-biased judgment, namely, decision threshold shifts that are implemented differently as a function of the motivational context. Consistent with previous research, the present findings suggest that the MOFC is not tied to the particular outcome of positively-biased evaluation, but rather that the MOFC is sensitive to the motivational significance of social decisions. In the context of threat, MOFC may predict liberal decision threshold shifts that increase positively-biased responses because these responses may be a valuable way to protect the self (Study 2). However, when threat is not explicitly heightened, MOFC may be sensitive to contextual factors such as restricted trait breadth that elicit conservative shifts that reduce the expression of positively-biased responses (Studies 1A and 1B). This interpretation parallels research showing that MOFC influences the way evidence for a perceptual judgment is evaluated as a function of motivational contexts that bias those judgments (Basten et al., 2010; Mulder et al., 2012; Scheibe et al., 2010; Summerfield & Koechlin, 2008, 2010). Research in perceptual decision-making characterizes bias as shifts in the starting point in the accumulation of evidence that favors a decision (Ratcliff & McKoon, 2008), and there is evidence that MOFC is sensitive to those shifts in starting points (Mulder et al., 2012; Scheibe et al., 2010).

While we are proposing that MOFC may support a common decision threshold shift function that is implemented differently depending on the motivational context, the present work cannot discount the possibility that MOFC supports different processes in different motivational contexts. For example, it may be that the same MOFC region supports one mechanism when threat is not explicitly heightened, and a different mechanism in the context of explicit threat, because self-evaluation motivations were not manipulated within the same participants. In order to directly associate MOFC activation with a common mechanism that up-regulates and down-regulates positively-biased responses as a function of motivational contexts, it will be important for future research to manipulate self-evaluation motivations within the same participants. For example, does MOFC predict accurate or consistent self-evaluations as well as positively-biased evaluations to the extent that people are motivated to achieve those outcomes in their self-evaluations (e.g., Brown, 2012; Trope, 1986; Swann, 1983; Sedikides et al., 2007)? Answering these questions has the potential to further constrain our understanding of whether positivity biases reflect multiple processes as a function of motivational contexts.

Although the present research suggests that MOFC may shift decision thresholds away from baseline positively-biased starting points as a function of the motivational context, questions remain about whether these baseline starting points exist, where in the brain they are represented, and how they may be updated as a function of different motivations. As mentioned above, there is evidence to suggest that positivity bias is prepotent or at least relies on relatively automatized processing (Alicke et al., 1995; Beer & Hughes, 2010; Beer, Chester, & Hughes, forthcoming; Hixon & Swann, 1983; Hughes

& Beer, unpublished data; Koole et al., 2001; Lench & Ditto, 2008; Paulhus et al., 1989; Swann et al., 1990). One avenue to test the presence of prepotent or baseline positively-biased starting points is to examine populations that decouple the direction of decision threshold shift and the direction of its impact on the expression of positively-biased responses. For example, people with depressed or low self-esteem tend to have baseline associations with self that are negative rather than positive (e.g., Koole et al., 2001; Swann & Read, 1981 and see Phillips, Hine, & Thorsteinsson, 2010 for a review) and can respond to self-esteem threats with increased self-deprecation (e.g., Vohs & Heatherton, 2004 and see vanDellen et al., 2011 for a review). In these populations, MOFC engagement may predict the combination of a shift towards a more liberal threshold underlying an increase in responses that are negatively-biased (rather than positively-biased) as a function of explicitly heightened threat.

Another avenue to begin to examine baseline positively-biased starting points may be to apply computational modeling techniques that provide measures that approximate these baseline starting points in self-evaluation. For example, "model based" reinforcement learning and Bayesian analysis provide a way to approximate internal models and prior beliefs that bias behavior, which may serve as proxies for baseline starting points in self-evaluation (e.g., Daw et al., 2011; Griffiths et al., 2010; Hampton et al., 2006). Bayesian models dictate how agents should update their beliefs in light of new information and the strength of the prior knowledge possessed by the agent. The extent to which people update or are resistant to updating these prior beliefs given new information may provide a method for measuring baseline starting points and their relation to the

neurobiology of positively-biased judgment. Similarly, drift diffusion models permit the measurement of shifts in starting points in evidence accumulation that may bias decisions (Mulder et al., 2012; Ratcliff & McKoon, 2008). The relation between variability in individuals' starting points and shifts in positively-biased responses as a function of the motivational context may provide additional insight into the existence of baseline starting points in positivity bias.

Role of Medial Prefrontal Cortex in Positivity Bias

While the MOFC may support a common mechanism in positively-biased judgment that is implemented differently as a function of the motivational context, the MPFC may support a distinct mechanism or a distinct neural interaction as a function of a heightened self-protection motivation. A consideration of the neuroanatomy and functions associated with MPFC may be useful to inform its role in positively-biased judgment.

NEUROANATOMY OF THE MPFC

The neuroanatomy of MPFC suggests possibilities about its interaction with neural regions related to positivity biases as well as the potential mechanisms it might support in positively-biased judgment. The MPFC, consisting of Bodmann areas 9 and 10, is neuroanatomically distinct from the MOFC region in BA 11 described above (Ongur & Price, 2000; Price & Drevets, 2010). First, the MPFC (BA 9 and 10) is robustly interconnected with other medial prefrontal cortical areas, such as the MOFC (BA

11)(Barbas et al., 1999; Price & Drevets, 2010). Second, evidence for memory-related input into the MPFC is suggested by its interconnection with memory-related regions such as the entorhinal cortex, parahippocampal cortex, hippocampal formation, posterior cingulate cortex (PCC), and lateral parietal lobe (Barbas et al., 1999; Price & Drevets, 2010). In fact, the MPFC and memory-related regions listed above form part of the default mode network of brain regions that are functionally as well as anatomically connected (Andrews-Hanna et al., 2010; Vincent et al., 2006). These default mode regions are active at rest and during tasks that encourage an internal focus of attention and exhibit a pattern of deactivation during certain goal-directed behaviors, potentially to suspend attention to internal information that may interfere with attention to the external environment (Andrews-Hanna et al., 2010; Gusnard & Raichle, 2001). One possibility is that the MPFC, via its interaction with the default mode network, may support access to internal information necessary for positivity biases in self-evaluations in the context of threat, and the MOFC, via its interaction with the MPFC, may modulate the expression of the accessed information. A more specific consideration of the distinct mechanism that may be supported by MPFC in positively-biased judgment is discussed below in relation to psychological processes associated with MPFC function.

FUNCTIONS OF THE MPFC

The MPFC is consistently recruited in a variety of tasks that require accessing internal representations of self as compared to a variety of semantic control conditions (e.g., Craik et al., 1999; Gillihan & Farah, 2005; Johnson et al., 2002; Kelley et al.,

2002). Previous research shows that MPFC activation is recruited when people evaluate the self-descriptiveness and certainty of personality traits (D'Argembeau et al., 2012; Fossati et al., 2003; Macrae et al., 2004; Moran et al., 2006; Ochsner et al., 2005), evaluate their preferences and attitudes (Cunningham et al., 2004; Jenkins et al., 2008; Mitchell et al., 2006; Zysset et al., 2002), evaluate or monitor their task performance (Beer et al., 2010; Bengtsson, et al., 2009), and evaluate their personality across time (D'Argembeau et al., 2008, 2009; Ersner-Hershfield et al., 2009; Tamir & Mitchell, 2011). Consistent with its relation to a network of memory-related regions, MPFC activation predicts subsequent memory for information that was processed in a self-referential manner (Macrae et al., 2004), and MPFC is engaged during the retrieval of information that was encoded in a self-referential manner (Fossati et al., 2004; Benoit et al., 2010). Taken together, evidence suggests that the MPFC is important for accessing and representing aspects of the self.

The relation between MPFC and self-processing is potentially due to its broader role in accessing and representing internally generated information in general, which encompasses self-information (Burgess et al., 2007; Christoff & Gabrieli, 2000; Christoff et al., 2004; Passingham et al., 2009). For example, MPFC activation is recruited by a number of disparate tasks that all require representations of internally generated information, such as recalling autobiographical memories (Cabeza & St. Jacques, 2007; Schacter, Addis, & Buckner, 2007), imagining future situations (*self-projection*: Buckner & Carroll, 2007; Gilbert & Wilson, 2007; Szpunar, Watson, & McDermott, 2007), engaging in spatial navigation (Hassabis & Maguire, 2007), mind-wandering and

stimulus-independent thought (Christoff et al., 2009; Mason et al., 2007), and thinking about other people (*mentalizing*: Blakemore et al., 2004; Harris et al. 2005; Krienen et al., 2010; Mitchell et al., 2006). The role of MPFC in accessing internally generated information is consistent with the observation that MPFC is part of the default mode network discussed above that is active at rest and during tasks that encourage attention to internally generated thoughts and feelings that may be necessary to form mental representations of self (Andrews-Hanna et al., 2010; Buckner & Carroll, 2007; Schachter et al., 2007; Spreng, Mar, & Kim, 2008). Taken together, these empirical observations support the notion that MPFC plays a role in internal representations of information about the self, and this process may play a distinct role in positively-biased evaluations in the context of threat.

The findings from the present set of studies contribute to this large literature on MPFC function and to ongoing discussions about how MPFC and MOFC may play distinct roles in self-processing and positivity biases in self-evaluation (e.g., Beer, 2007; D'Argembeau et al., 2012; Northoff & Bermpohl, 2004; Ochsner et al., 2005). One possibility is that the MPFC may support access to more certain aspects of self while the MOFC evaluates whether the accessed information should be expressed in judgment based on contextual or motivational goals. People respond to threat by drawing on portions of their mental representations of self in addition to or instead of emotion-regulation processes such as inhibition or reappraisal of the threat stimulus itself (Baumeister & Jones, 1978; Wood et al., 1999). In particular, people respond to threatening feedback by accessing core aspects of their self-concept (Aronson et al.,

1995; Dodgson & Wood, 1998; Steele et al., 1993; vanDellen et al., 2011; Wood et al., 1999) and increase their influence on subsequent self-evaluations to the extent that they find it believable or defensible (Kunda, 1990; Sedikides et al., 2002). This possibility is consistent with the research described above that posits a role for MPFC in mental representations of self, as well as recent research demonstrating that MPFC differentiates self-judgments of personality that are more certain from those that are more uncertain (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. Therefore, MPFC's role in accessing certainly held aspects of the self-concept might be particularly important in the context of threat. However, it remains unknown whether MPFC activation reflects access to core aspects of the self when threat is explicitly heightened, and whether these core aspects increase positivity bias in self-evaluations. Studies that examine evaluations of core and non-core aspects of self under conditions that vary in the extent to which threat is explicitly heightened may begin to address this possibility.

The expression of self-information accessed by the MPFC may be influenced by decision threshold shifts supported by MOFC activation, a possibility that is consistent with the neuroanatomical connectivity between MPFC and MOFC. This perspective may explain why MOFC activation differentiates judgments about traits that are deemed important to possess compared to traits deemed unimportant to possess (D'Argembeau et al., 2012). 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). In the context of threat, MOFC activation may shift decision thresholds in order to liberally define traits in ways that allow the self to appear special in order to compensate for the threat. While this suggestion is consistent with the MPFC and MOFC connectivity described above (Amodio & Frith, 2006; Barbas et al., 1999; Price & Drevets, 2010), more research is needed to fully understand whether MOFC is a region that modulates the expression of self-representations mediated by MPFC function.

Role of Amygdala and Insula in Positivity Bias

While the roles of the amygdala and insula in positivity bias remain underdetermined, one possibility is that amygdala and insula may project important affective signals to the MOFC and MPFC in the context of explicit threat. A host of research and recent meta-analyses have shown that the amygdala and insula are important structures for processing the affective properties of information, and emotion's influence on cognitive processes such as attention, memory, and decision-making (Kober et al., 2008; Phan et al., 2002; Phelps, 2006; Singer, Critchley, & Preuschoff, 2009). The amygdala is believed to play a role in processing motivationally relevant stimuli as a function of the situation (Anderson & Phelps, 2001; Cunningham et al., 2008; Pessoa et al., 2006; Todorov, Baron, & Oosterhof, 2008), which may explain the classic association between amygdala activation and threat-related processing (LeDoux, 2000; Whalen, 1998). For example, the processing goals of an observer or the chronic motivational styles that observers use to deal with affective information (e.g., neuroticism) modulate

whether amygdala signals rewarding information, threatening information, or both (Cunningham et al., 2008, 2010). Therefore, the amygdala may signal information that is motivationally relevant as a function of the situation, and modulate attention, memory, and decision-making to deal with motivationally salient events.

Similarly, the insula is similarly implicated in affective processing (Nitschke et al., 2006; Ploghaus et al., 1999; Phan et al., 2002; Kober et al., 2008), with increased insular sensitivity to affective information in individuals with high trait anxiety and anxiety disorders (Etkin & Wager, 2007; Stein et al., 2007). Recent evidence suggests that the insula may continuously monitor threat levels in the environment, with anxious individuals showing increased insula activation during threat monitoring (Somerville, Whalen, & Kelley, 2010b). Moreover, the insula is important for representing the visceral feeling states associated with emotional experiences (Craig, 2002; Singer et al., 2009; Critchley, 2005). For example, insula activity is associated with receiving unfair offers and subsequently rejecting those unfair offers (Sanfey et al., 2003), and with the experience of social pain and empathy during social interaction tasks (Eisenberger et al., 2003, 2011; Hein et al., 2010; Kross et al., 2007; Singer et al., 2006). Therefore, by representing current feeling states, the insula may serve as a benchmark for interactions with the environment (Critchley, 2005).

Taken together, previous research is consistent with the hypothesis that amygdala and insula may project current motivational and affective signals to the MOFC and MPFC in order to guide the access and evaluation of self-related information in order to compensate for threat. This hypothesis is supported by evidence that both amygdala and

insula are interconnected with the MOFC (Cavada et al., 2000; Ongur & Price, 2000). The connectivity between these regions and the MOFC are important for emotional and reinforcement learning (Baxter & Murray, 2002; Hampton et al., 2007; Holland & Gallagher, 2004; Milad & Quirk, 2002; Phelps et al., 2004) and their interactions with the hippocampus, a memory region that is interconnected with both MPFC and MOFC, have been shown to modulate emotional and autobiographical memory (Adolphs et al., 2005; Dolcos et al., 2005; Sharot et al., 2007b). While amygdala and insula activation did not predict increases in positively-biased responses as a function of threat (Study 2), activation in these regions might modulate activity in MOFC or MPFC regions associated with positively-biased responses as a function of threat. Therefore, future research may find distinct connectivity patterns between amygdala, insula and medial cortical regions as a function of threat.

A second possibility is that amygdala and insula are related to within-subject variability in positively-biased responses, rather than between-subject variability as measured in the present work. For example, amygdala and insula activation are often associated with processing affective information and feeling states in a continuous trial-by-trial within-subjects manner (e.g., Canli et al., 2000; Somerville et al., 2010b). Therefore, future studies that use tasks that allow for a trial-by-trial mapping of amygdala and insula activation and positively-biased responses as a function of threat may uncover an association between amygdala and insula activation and within-subject variability in positivity bias as a function of threat.

Limitations

One primary limitation of the present work is the heavy focus on better-than-average judgments as the operationalization of positivity bias (Studies 1A, 1B, and 2). Better-than-average judgments were chosen as a primary operationalization of positivity bias for theoretically motivated reasons. As mentioned above, some researchers have questioned whether positivity biases reflect a self-protection motivation because of a relative lack of demonstrations that threat increases better-than-average responses (Chambers & Windschitl, 2004; but see Brown, 2012; Vohs & Heatherton, 2004). Therefore, examining the neural mechanisms of better-than-average responses in situations with threat and without heightened threat was particularly useful to determine whether positivity biases reflect a single phenomenon or multiple distinct phenomena as a function of motivational context. However, recent research has called into question the extent to which better-than-average judgments reflect a biased judgment or a self-serving judgment (Giladi & Klar, 2002; Klar & Giladi, 1999; also see Chambers & Windschitl, 2004). First, research has shown that when people are asked to evaluate how happy they are compared to others in their peer group, people base their social comparative judgments largely on their own level of happiness. Specifically, social comparative judgments of happiness were strongly related to absolute judgments of one's own happiness, but unrelated to absolute judgments of other people's happiness (Klar & Giladi, 1999). Therefore, the definition of positivity bias in social comparisons as the degree of deviation from the average peer may reflect a self-evaluation without

consideration of the referent group, which suggests that better-than-average judgments may be a problematic measure of a positivity bias. Second, research has shown that people evaluate randomly selected individuals of a group (even nonsocial objects such as soaps and songs) more favorably than other members of that group (Klar & Giladi, 1997; Giladi & Klar, 2002). The arbitrary nature of the selection of the target for comparison raises the possibility that better-than-average judgments reflect a non-selective superiority bias rather than a more specific self-serving positivity bias. Giladi and Klar (2002) present a local-comparisons-general-standards (LOGE) approach to explain the better-than-average effect. Specifically, the LOGE approach posits that when people compare one target member of a group to other members of that group, people fail to use appropriate standards specific to the comparison group and instead use more general standards involving members from outside the comparison group. Therefore, better-than-average biases may represent a more general class of judgment biases that are not self-specific. It is important to note that Study 3 and previous research (Beer et al., 2006; Beer et al., 2010; Krusemark et al., 2008) used very different operationalizations of positivity bias (e.g., overclaiming bias, overconfidence bias) and different manipulations to generate variance in positively-biased responses and found convergent evidence for the involvement of a core set of neural regions (OFC, MPFC, dACC). Future studies should use a variety of methods to operationalize positivity bias in order to find convergent evidence for the neural associations of positivity bias in self-evaluations.

Second, while the better-than-average task and behavioral manipulations used in the present research provide a measure of between-subject differences in positivity bias,

they are limited in their ability to provide trial-by-trial, or within-subject differences in positivity bias. First, the better-than-average effect does not provide a measure of positivity bias at the trial level of analysis: At the trial level, it is impossible to tell apart individuals who make positively-biased evaluations from those that are exceptional on that trait. Second, the threat and accountability manipulations provide measures of shifts in positively-biased evaluations as a function of motivational contexts between blocks, but do not provide a way to examine the neural regions that may be linked to positively-biased responses at the trial level of analysis. Therefore, the amygdala, insula, and other regions that were engaged in conditions of positively-biased evaluation but were not modulated by between-subject variability in positively-biased evaluation may be modulating behavioral responses at the trial level of analysis. One previous study included a trial-by-trial measure of bias (overconfidence in task performance) and found convergent evidence that MOFC tracks trial-by-trial variability in confidence estimates (Beer et al., 2010). Future research that examines positivity bias on a trial-by-trial basis may be helpful for providing convergent evidence for the proposed neural model of positively-biased judgment as well as help extend the neural model by incorporating relationships that might have been missed with the current approach (e.g., Beer et al., 2010).

A final limitation is that the behavioral manipulations used may have contributed to decreased statistical power. For example, extensive pilot testing revealed that the threat manipulation used in Study 2 might have a diminished effect on people's behavioral responses after too many exposures to the threatening feedback. This limited the number

of blocks that could be successfully implemented in the fMRI study. While the blocked nature of the task may partially alleviate the decreased statistical power associated with a limited number of repetitions, it does not discount the possibility that explicit threat may have engaged additional neural regions that we did not have the power to detect. In addition, neural regions that were associated with positively-biased evaluations in some but not all of the studies, such as LOFC (Studies 1A and 1B), MPFC (Study 2), and dACC (Study 1A), may support common mechanisms across different motivational contexts much like MOFC, but they were not detected across all studies due to lack of power. Therefore, future research that examines the mechanisms supported by additional neural regions associated with positivity bias will help shed light on whether positivity biases reflect multiple distinct processes as a function of whether self-protection motivation is heightened.

Conclusions

How do people make positively-biased evaluations of their personality, knowledge, and behavior in different contexts? Do positively-biased evaluations represent a single phenomenon or multiple distinct phenomena depending on contextual motivators? The experiments presented here attempt to deepen our understanding of how positively-biased evaluations occur as a function of different motivational contexts. Peering inside the brain to understand the neural underpinnings of positivity biases and how these neural systems are influenced by different motivations has the potential to inform our understanding of how positively-biased evaluations are accomplished, a question that has been problematic to address with behavioral measures alone. On a broader level, understanding how positivity biases occur could have far reaching real-world implications. For example, a deeper understanding of how positivity biases occur may be helpful for providing methods to adjust some of the documented maladaptive effects of positivity biases on health behaviors, educational outcomes, and in the work place (Dunning et al., 2004). Similarly, future research along these lines may shed new light on why different forms of flawed self-evaluation arise in clinical populations characterized by neurological impairments. For example, mood disorders and substance abuse are associated with impairments in medial cortical areas and anatomically related limbic structures (Price & Drevets, 2010; Volkow et al., 1991) and are characterized by impaired self-insight (Aleman et al., 2006; Alloy & Ahrens, 1987). Understanding the relationship between neural impairment and flawed self-assessment may facilitate the

development of interventional therapies by pinpointing the processes and patients who are likely to benefit from such interventions.

References

Adolphs R., Tranel D., Buchanan T.W. (2005). Amygdala damage impairs emotional memory for gist but not details of complex stimuli. *Nat. Neurosci*. 8, 512–18

Alicke, M.D. (1985). Global self-evaluation as determined by the desirability and controllability of trait adjectives. *Journal of Personality and Social Psychology, 49,* 1621-1630.

Alicke, M. D., Klotz, M. L., Breitenbecher, D. L., Yurak, T. J., & Vredenburg, D. S. (1995). Personal Contact, Individuation, and the Better-Than-Average Effect. *Journal of Personality and Social Psychology, 68*,804-825.

Alicke, M.D., & Govorun, O. (2005). The better-than-average effect. In M.D. Alicke, D.A. Dunning, & J.I. Krueger (Eds.), *Studies in self and identity* (pp. 85-106). New York: Psychology Press.

Alloy, L. B., & Ahrens, A. H. (1987). Depression and pessimism for the future: Biased use of statistically relevant information in predictions for self versus others. *Journal of Personality and Social Psychology, 52*(2), 366-378.

Amaral, D.G., & Price, J.L. (1984). Amygdalo-cortical projectsion in monkey (Macaca fascicularis). *Journal of Computational Neurology, 230,* 465-496.

Amodio, D.M., & Frith, C.D. (2006). Meeting of minds: The medial frontal cortex and social cognition. *Nature Reviews Neuroscience, 7,* 268-277.

Anderson NH. (1968). Likability ratings of 555 personality trait words. *J. Pers. Soc. Psych. 9:* 272-279.

Anderson, C., Srivastava, S., Beer, J.S., Spataro, S.E., & Chatman, J.A. (2006). Knowing your place: Self-perceptions of status in face-to-face groups. *Journal of Personality and Social Psychology, 91,* 1094-1110.

Anderson, A. K., & Phelps, E. A. (2001). Lesions of the human amygdala impair enhanced perception of emotionally salient events. *Nature, 411,* 305–309.

Andrews-Hanna, J.R., Reidler, J.S., Sepulcre, J., Poulin, R., Buckner, R.L. (2010). Functional-anatomic fractionation of the brain's default network. *Neuron, 65,* 550-562.

Arana, F.S., Parkinson, J.A., Hinton, E., Holland, A.J., Owen, A.M., Roberts, A.C. (2003). Dissociable Contributions of the Human Amygdala and Orbitofrontal

Cortex to Incentive Motivation and Goal Selection. *Journal of Neuroscience, 23*, 9632-9638

Aronson, J., Blanton, H., Cooper, J. (1995). From dissonance to disidentification: Selectivity in the self-affirmation process. *Journal of Personality and Social Psychology, 68,* 986-996.

Aronson, E., & Carlsmith, J.M. (1968). Experimentation in social psychology. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology* (2nd ed., Vol. 2). New York: Addison-Wesley.

Balleine, B.W., & O'Doherty, J.P. (2010). Human and rodent homologies in action control: Corticostriatal determinants of goal-directed and habitual action. *Neuropsychopharmacology, 35,* 48-69.

Barbas, H., Ghashghaei, H., Dombrowski, S.M., Rempel-Clower, N.L. (1999). Medial prefrontal cortices are unified by common connections with superior temporal cortices and distinguishes input from memory-related areas in rhesus monkey. *Journal of Computational Neurology, 410,* 343-367.

Basten, U., Biele, G., Heekeren, H.R., Fiebach, C.J. (2010). How the brain integrates costs and benefits during decision making. *Proceedings of the National Academy of Sciences, 107*, 21767-21772.

Baumeister, R.F. & Jones, E. E. (1978). When self-presentation is constrained by the target's knowledge: Consequence and compensation. *Journal of Personality and Social Psychology, 36,* 608-618.

Baumeister, R.F., Heatherton, T.F., Tice, D.M. (1993). When ego threats lead to self-regulation failure: Negative consequences of high self-esteem. *Journal of Personality and Social Psychology, 64,* 141-156.

Baumeister, R.F., DeWall, C.N., Ciarocco, N.J., & Twenge, J.M. (2005). Social exclusion impairs self-regulation. *Journal of Personality and Social Psychology, 88,* 589-604.

Baxter, M.G., & Murray, E.A. (2002). The amygdala and reward. *Nature Reviews Neuroscience, 3,* 563-573.

Bechara, A., Tranel, D., Damasio, H. (2000). Characterization of the decision-making deficit of patients with ventromedial prefrontal cortex lesions. *Brain, 123,* 2189-2202.

Beer, J.S., John, O.P., Scabini, D., & Knight, R.T. (2006). Orbitofrontal cortex and social behavior: Integrating self-monitoring and emotion-cognition interactions. *Journal of Cognitive Neuroscience, 18,* 871-888.

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 comparisons and the "above average" effect. *NeuroImage, 49,* 2671-2679.

Beer, J. S., Lombardo, M. V., & Bhanji, J. P. (2010). Roles of medial prefrontal cortex and orbitofrontal cortex in self-evaluation. *Journal of Cognitive Neuroscience.* 22, 2108-2119.

Beer, J.S., Chester, D.S., & Hughes, B.L. (forthcoming). Worse-than-average, but not if you say so or I'm busy: The effects of self-esteem threat and cognitive load on social comparisons. Manuscript submitted for publication.

Beer, J. S., Shimamura, A. P., & Knight, R. T. (2004). Frontal lobe contributions to executive control of cognitive and social behavior. In M. S. Gazzaniga (Ed.), The Cognitive Neurosciences III (pp. 1091-1104). Cambridge: MIT Press.

Behrens, T.E.J., Hunt, L.T., Woolrich, M.W., & Rushworth, M.F.S. (2008). Associative learning of social value. *Nature, 456,* 245-249.

Bengtsson, S.L., Lau, H.C., Passingham, R.E. (2009). Motivation to do well enhances responses to errors and self-monitoring. *Cerebral Cortex, 19,* 797-804.

Benoit, R.G., Gilbert, S.J., Volle, E., Burgess, P.W. (2010). When I think about me and simulate you: Medial rostral prefrontal cortex and self-referential processes. *NeuroImage, 50,* 1340-1349.

Bing, M.N., Kluemper, D., Davison, H., Taylor, S., & Novicevic, M. (2011). Overclaiming as a measure of faking. *Organizational Behavior and Human Decision Processes, 11,* 148-162.

Bhanji, J. P., Beer, J. S. (in press). Taking a different perspective: Mindset influences neural regions that represent value and choice. *Soc. Cogn. Aff. Neurosci.* Advance Online Access.

Blaine, B., Crocker, C. (1993). Self-esteem and self-serving biases in reactions to positive and negative events: An integrative review. In R.F. Baumeister (Ed.), *Self-esteem: The puzzle of low self-regard* (pp. 21-36). New York, NY: Plenum.

Blakemore, S.J., Winston, J., Frith, U. (2004). Social cognitive neuroscience: Where are we heading? *Trends in Cognitive Sciences, 8,* 216-222.

Blanton, H. Axsom, D., McClive, K.P., & Price, S. (2001). Pessimistic bias in comparative evaluations: A case of perceived vulnerability to the effects of negative life events. *Personality and Social Psychology Bulletin, 27*, 1627-1636.

Botvinick, M.M., Cohen, J.D., & Carter, C.S. (2004). Conflict monitoring and anterior cingulate cortex: An update. *Trends in Cognitive Science, 8,* 539-546.

Bradley, G.W. (1978). Self-serving biases in the attribution process: A reexamination of the fact or fiction question. *Journal of Personality and Social Psychology, 36,* 56-71.

Brett, M., Anton, J.L., Valabregue, R., Poline, J.B. (2002). Region of interest analysis using an SPM toolbox. *NeuroImage, 16*, 1140-1141.

Brown, J. D. (1986). Evaluations of self and others: Self-enhancement biases in social judgment. *Social Cognition*, *4*, 353–376.

Brown, J.D. (2012). Understanding the better than average effect: Motives (still) matter. *Personality and Social Psychology Bulletin, 38,* 209-219.

Brown, J.D., Dutton, K.A. (1995). The thrill of victory, the complexity of defeat: Self-esteem and people's emotional reactions to success and failure. *Journal of Personality and Social Psychology, 68,* 712-722.

Brown, J.D., Smart, S.A. (1991). The self and social conduct: Linking self-representations to prosocial behavior. *Journal of Personality and Social Psychology, 60,* 368-375.

Buckner, R.L., & Carroll, D.C. (2007). Self-projection and the brain. *Trends in Cognitive Sciences, 11,* 49-57.

Burgess, P.W., Dumonthiel, I., Gilbert, S.J. (2007). The gateway hypothesis of rostral prefrontal cortex (area 10) function. *Trends in Cognitive Sciences, 11,* 290-298.

Buss D.M., Craik K.H. (1983). The act frequency approach to personality. *Psych. Rev. 90*: 105-126.

Buunk BP, Van Yperen NW. (1991). Referential comparisons, relational comparisons, and exchange orientation: Their relation to marital satisfaction. *Pers. and Soc. Psych. Bull. 17:* 709.

Cabeza, R., & St. Jacques, P. (2007). Functional neuroimaging of autobiographical memory. *Trends in Cognitive Sciences, 11,* 219-227.

Campbell, J. D. (1986). Similarity and uniqueness: The effects of attribute, type, relevance, and individual differences in self esteem and depression. *Journal of Personality and Social Psychology, 50*, 281-294.

Campbell, J.D. (1990). Self-esteem and clarity of the self-concept. *Journal of Personality and Social Psychology, 59,* 538-549.

Campbell, W.K., Sedikides, C. (1999). Self-threat magnifies the self-serving bias: A meta-analytic integration. *Review of General Psychology, 3,* 23-43.

Canli, T., Zhao, Z., Brewer, J., Gabrieli, J.D.E., & Cahill, L. (2000). Event-related activation in the human amygdala associates with later memory for individual emotional experience. *Journal of Neuroscience, 20,* 1–5.

Cavada, C., Company, T., Tejedor, J., Cruz-Rizzolo, R.J., Reinoso-Suarez, F. (2000). The anatomical connections of the macaque monkey orbitofrontal cortex: A review. *Cerebral Cortex, 10,* 220-242.

Chambers, J.R., Windschitl, P.D. (2004). Biases in social comparative judgments: The role of nonmotivated factors in above-average and comparative-optimism effects. *Psychological Bulletin, 130,* 813-838.

Chambers, J. R., Windschitl, P. D., & Suls, J. (2003). Egocentrism, Event Frequency, and Comparative Optimism: When what Happens Frequently is" More Likely to Happen to Me". *Personality and Social Psychology Bulletin, 29*(11), 1343.

Christoff, K., Ream, J.M., Gabrieli, J.D.E. (2004). Neural basis of spontaneous thought processes. *Cortex, 40,* 623-630.

Christoff, K., & Gabrieli, J.D.E. (2000). The frontopolar cortex and human cognition: Evidence for a rostrocaudal hierarchical organization within the human prefrontal cortex. *Psychobiology, 28*, 168-186.

Christoff, K., Gordon, A.M., Smallwood, J., Smith, R., Schooler, J.W. (2009). Experience sampling during fMRI reveals default network and executive system

contributions to mind wandering. *Proceedings of the National Academy of Sciences, 106,* 8719-8724.

Cacioppo, J. T., & Bernston, G. G. (1992). Social psychological contributions to the Decade of the Brain: Doctrine of multilevel analysis. *American Psychologist, 47,* 1019-1028.

Colvin, C. R., Block, J., & Funder, D. C. (1995). Overly positive self-evaluations and personality: Negative implications for mental health. *Journal of Personality and Social Psychology, 68*(6), 1152-1162.

Cooney, R. E., Joormann, J., Eugene, F., Dennis, E. L., Gotlib, I.H. (2010). Neural correlates of rumination in depression. *Cog. Aff. Beh. Neuro., 10*, 470-478.

Cooper, A.C., Woo, C.Y., & Dunkelberg, W.C. (1988). Entrepreneurs' perceived chances for success. *Journal of Buisiness Venturing, 3,* 97-108.

Craig, A.D. (2002). Opinion: How do you feel? Interoception: The sense of the physiological condition of the body. *Nature Reviews Neuroscience, 3,* 655-666.

Craik, F.I.M., Moroz, T.M., Moscovitch, M., Stuss, D.T., Winocur, G., Tulving, E., Kapur, S. (1999). In search of the self: A positron emission tomography study. *Psychological Science, 10*, 26-34

Critcher, C.R., & Dunning, D. (2009). How chronic self-views influence (and mislead) self-assessments of task performance: Self-views shape bottom-up experiences with the task. *Journal of Personality and Social Psychology, 97,* 931-945.

Critchley, H.D. (2005). Neural mechanisms of autonomic, affective and cognitive integration. *Journal of Comparative Neurology, 493,* 154-166.

Crocker, J., & Park, L.E. (2004). The costly pursuit of self-esteem. *Psychological Bulletin, 130,* 392-414.

Cross, P. (1977). Not can but will college teaching be improved. *New Directions for Higher Education, 17,* 1-15.

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

Cunningham, W. A., Raye, C. L., & Johnson, M. K. (2004). Implicit and explicit evaluation: fMRI correlates of valence, emotional intensity, and control in the processing of attitudes. *Journal of Cognitive Neuroscience, 16,* 1717–1729.

Cunningham, W.A., Arbuckle, N.L., Jahn, A., Mowrer, S.M., Abdulhalil, A.M. (2010). Aspects of neuroticism and the amygdala: Chronic tuning from motivational styles. *Neuropsychologia, 48,* 3399-3404.

D'Argembeau, A., Jedidi, H., Balteau, E., Bahri, M., Phillips, C., Salmon. E. (2012). Valuing one's self: Medial prefrontal involvement in epistemic and emotive investments in self-views. *Cerebral Cortex, 22,* 659-667.

D'Argembeau, A., Feyers, D., Majerus, S., Collette, F., Van der Linden, M., Maquet, P., Salmon,E. (2008). Self-reflection across time: Cortical midline structures differentiate between present and past selves. *Social Cognitive Affective Neuroscience, 3,* 244-252.

D'Argembeau, A., Stawarczyk, D., Majerus, S., Collette, F., Van der Linden, M., Feyers, D., Maquet, P., Salmon, E. (2009). The neural basis of personal goal processing when envisioning future events. *Journal of Cognitive Neuroscience.*

Daw, N.D., O'Doherty, J.P., Dayan, P., Seymour, B., & Dolan, R.J. (2006). Cortical substrates for exploratory decisions in humans. *Nature, 441,* 876-879.

Daw, N.D., Gershman, S.J., Seymour, B., Dayan, P., Dolan, R.J. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron, 69,* 1204-1215.

De Araujo, I.E., Rolls, E.T., Velazco, M.I., Margot, C., Cayeux, I. (2005). Cognitive modulation of olfactory processing. *Neuron, 46,* 671-679.

De Martino, B., Kumaran, D., Seymour, B., and Dolan, R. J. (2006). Frames, biases, and rational decision-making in the human brain. *Science, 313,* 684-687.

Dunning, D., Heath, C., Suls, J.M. (2004). Flawed self-assessment: Implications for health, education and the workplace. *Psychological Science in the Public Interest, 5,* 69-106.

Dunning, D. (1995). Trait importance and modifiability as factors influencing self-assessment and self-enhancement motives. *Pers. Soc. Psych. Bull. 21*, 1297-1306.

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. *J. Pers. Soc. Psych. 57*, 1082-1090.

Dunning, D., Beauregard, K.S. (2000). Regulating impressions of others to affirm images of the self. *Social Cognition, 18,* 198-222.

Dodgson, P.G., Wood, J.V. (1998). Self-esteem and the cognitive accessibility of strengths and weaknesses after failure. *Journal of Personality and Social Psychology, 75,* 178-197.

Dolcos, F., LaBar, K.S., & Cabeza, R. (2005). Remembering one year later: Role of the amygdala and the medial temporal lobe memory system in retrieving emotional memories. *Proceedings of the National Academy of Sciences, 102,* 2626-2631.

Donaldson, D.I., Petersen, S.E., Ollinger, J.M., Buckner, R.L. (2001). Dissociating state and item components of recognition memory using fMRI. *NeuroImage, 13*, 129-142.

Eisenberger, N.I., Lieberman, M.D., & Williams, K.D. (2003). Does rejection hurt? An fMRI study of social exclusion, *Science, 302,* 290-292.

Eisenberger, N.I., Inagaki, T.K., Muscatell, K.A., Haltom, K.E.B., Leary, M.R. (2012). The neural sociometer: Brain mechanisms underlying state self-esteem. *Journal of Cognitive Neuroscience.*

Ersner-Hershfield, H., Wimmer, G.E., Knutson, B. (2009). Saving for the future self: Neural measures of future self-continuity predict temporal discounting. *Social Cognitive Affective Neuroscience, 4,* 85-92.

Etkin, A., & Wager, T.D. (2007). Functional neuroimaging of anxiety: A meta-analysis of emotional processing in PTSD, social anxiety disorder, and specific phobia. *American Journal of Psychiatry, 164,* 1476-1488.

Fein, S., & Spencer, S.J. (1997). Prejudice as self-image maintenance: Affirming the self through derogating others. *Journal of Personality and Social Psychology, 73,* 31-44.

Fellows, L.K., Farah, M.J. (2003). Ventromedial frontal cortex mediates affective shifting in humans: Evidence from a reversal learning paradigm. *Brain*, *126*, 1830-1837.

Fellows, L.K., & Farah, M.J. (2007). The Role of Ventromedial Prefrontal Cortex in Decision Making: Judgment under Uncertainty or Judgment Per Se? *Cerebral Cortex, 17*, 2669-2674.

Fischoff, B., Slovic, P., & Lichtenstein, S. (1977). Knowing with certainty: The appropriateness of extreme confidence. *Journal of Experimental Psychology: Human Perception and Performance, 3,* 552-564.

Fiske, S. T, & Taylor, S. E. (1991). Social cognition (2nd ed.). New YorkMcGraw-Hill.

Fossati, P., Hevenor, S.J., Graham, S.J., Grady, C., Keightley, M.L., Craik, F., & Mayberg, H. (2003). In search of the emotional self: An fMRI study using positive and negative emotional words. *American Journal of Psychiatry, 160,* 1938-1945.

Fuster, J.M. (2001). The prefrontal cortex-An update. *Neuron, 30,* 319-333.

Gagne F.M., Lydon J.E. (2001). Mind-set and close relationships: When bias leads to (in) accurate predictions. *J. Pers. Soc. Psych. 81:* 85-96.

Gardner, W.L., Pickett, C.L., & Brewer, M.B. (2000). Social exclusion and selective memory: How the need to belong influences memory

Giladi, E.E., & Klar, Y. (2002). When standards are wide of the mark: Nonselective superiority and inferiority biases in comparative judgments of objects and concepts. *Journal of Experimental Psychology: General, 131,* 538-551.

Gilbert, D.T., & Wilson, T.D. (2007). Prospection: Experiencing the future. *Science, 317*, 1351-1354.

Gillihan, S.J., & Farah, M.J. (2005). Is self special? A critical review of evidence from experimental psychology and cognitive neuroscience. *Psychological Bulletin, 131,* 76-97.

Glascher, J., Hampton, A.N., & O'Doherty, J.P. (2009). Determining a role for ventromedial prefrontal cortex in encoding action-based value signals during reward-related decision making. *Cerebral Cortex, 19,* 483-495.

Goldman-Rakic, P.S. (1987). Circuitry of primate prefrontal cortex and regulation of behavior by representational knowledge. *Handbook of Physiology, 5,* 373-417.

Gosling, S. D., John, O. P., Craik, K. H., & Robins, R. W. (1998). Do people know how they behave? Self-reported act frequencies compared with on-line codings by observers. *Journal of Personality and Social Psychology, 74*, 1337–1349.

Gottfried, J.A., O'Doherty, J., & Dolan, R.J. (2003). Encoding predictive reward value in humanamygdala and orbitofrontal cortex. *Science, 301,* 1104-1107

Gramzow, R.H., & Willard, G. (2006). Exaggerating current and past performance: Motivated self-enhancement versus reconstructive memory. *Personality and Social Psychology Bulletin, 32,* 1114-1125.

Gray-Little, B., Williams, V.S.L., & Hancock, T. D. (1997). An item response theory analysis of the Rosenberg Self-esteem scale. *Personality and Social Psychology Bulletin*, 23, 443-451.

Green, D.M., Swets, J.A., 1966. Signal detection theory and psychophysics. Oxford: Wiley.

Greenberg, J., Pyszczynski, T. (1985). Compensatory self-inflation: A response to threat to self-regard of public failure. *Journal of Personality and Social Psychology, 49,* 273-280.

Griffiths, T.L., Chater, N., Kemp, C., Perfors, A., & Tenenbaum, J.B. (2010). Probabilistic models of cognition: Exploring representations and inductive biases. *Trends in Cognitive Sciences, 14,* 357-364.

Grinband, J., Savitskaya, J., Wager, T.D., Teichert, T., Ferrera, V.P., & Hirsch, J. (2011). The dorsal medial frontal cortex is sensitive to time on task, not response conflict or error likelihood. *NeuroImage, 57,* 303-311.

Gusnard, D.A., & Raichle, M.E. (2001). Searching for a baseline: Functional imaging and the resting human brain. *Nature Reviews Neuroscience, 2,* 685-694.

Hampson, S.E., John, O.J., Goldberg, L.R., (1986). Category breadth and hierarchical structure in personality: studies of asymmetries in judgments of trait implications. *J. Pers. Soc. Psych. 51*, 37–54.

Hampton, A.N., Bossaerts, P., & O'Doherty, J.P. (2006). The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans. *Journal of Neuroscience, 26,* 8360-8367.

Hampton, A.N., Adolphs, R., Tyszka, M.J., O'Doherty, J.P. (2007). Contributions of the amygdala to reward expectancy and choice signals in the human prefrontal cortex. *Neuron, 55,* 545-555.

Hare, T.A., Malmaud, J., Rangel, A. (2011). Focusing attention on the health aspects of foods changes value signals in vmPFC and improves dietary choice. *J. Neurosci., 31*, 11077-11087.

Harris, L.T., Todorov, A., Fiske, S.T. (2005). Attributions on the brain: Neuroimaging dispositional inferences, beyond theory of mind. *NeuroImage, 28,* 763-769.

Hassabis, D., & Maguire, E.A. (2007). Deconstructing episodic memory with construction. *Trends in Cognitive Sciences, 11,* 299-306.

Hein, G., Silani, G., Preuschoff, K., Batson, D.D., & Singer, T. (2010). Neural responses to ingroup and outgroup members' suffering predict individual differences in costly helping. *Neuron, 68,* 149-160.

Henson, R. (2006). Forward inference using functional neuroimaging: Dissociations versus associations. *Trends in Cognitive Sciences, 10,* 64-69.

Hixon, J.G., & Swann, W.B., Jr. (1993). When does introspection bear fruit? Self-reflection, self-insight, and interpersonal choices. *Journal of Personality and Social Psychology, 64,* 35-43.

Holland, P.C., & Gallagher, M. (2004). Amygdala-frontal interactions and reward expectancy. *Current Opinions in Neurobiology, 14,* 148-155.

Horton, R.S., & Sedikides, C. (2009). Narcissistic responding to ego threat: When the status of the evaluator matters. *Journal of Personality, 77*, 1493-1526.

Hughes, B. L. & Beer, J.S. (in press-a). Orbitofrontal cortex and anterior cingulate cortex are modulated by motivated social cognition. *Cerebral Cortex,* Advance Online Access.

Hughes, B. L. & Beer, J.S. (in press-b). Medial orbitofrontal cortex is associated with shifting decision thresholds in self-serving cognition. *NeuroImage.*

Hughes, B.L., & Beer, J.S. (forthcoming). Protecting the self: The effect of social evaluative threat on neural representations of self. Manuscript submitted for publication.

Hughes, B.L., & Beer, J.S. (unpublished data). Positively biased self-evaluation: Working

hard or hardly working?

Izquierdo, A., Suda, R.K., Murray, E.A. (2004). Bilateral orbital prefrontal cortex lesions in rhesus monkeys disrupt choices guided by both reward value and reward contingency. *Journal of Neuroscience, 24,* 7540-7548.

Jenkins, A. C., Macrae, C. N., & Mitchell, J. P. (2008). Repetition suppression of ventromedial prefrontal activity during judgments of self and others. *Proceedings of the National Academy of Sciences, 105(11)*, 4507-4512.

Johnson, S. C., Baxter, L. C., Wilder, L. S., Pipe, J. G., Heiserman, J. E., and Prigatano, G. P. (2002). Neural correlates of self-reflection. *Brain, 125*, 1808-1814.

Jones, E.E., & Nisbett, R.E. (1971). *The actor and the observer: Divergent perceptions of the causes of behavior.* Morristown, NJ: General Learning Press.

Jones, B., & Mishkin, M. (1972). Limbic lesions and the problem of stimulus-reinforcement associations. *Experimental Neurology, 36,* 362-377.

Josephs, R.A., Markus, H.R., Tafarodi, R.W. (1992). Gender and self-esteem. *Journal of Personality and Social Psychology, 63,* 391-402.

Kelley, W. M., Macrae, C. N., Wyland, C. L., Caglar, S., Inati, S., and Heatherton, T. F. (2002). Finding the self? An event-related fMRI study. *J Cog Neuro*, *14,* 785-794.

Kenny DA, Acitelli LK. (2001). Accuracy and bias in the perception of the partner in a close relationship. *J. Pers. Soc. Psych. 80:* 439-448.

Kirby DM, Gardner RC. (1972). Ethnic stereotypes: Norms on 208 words typically used in their assessment. *Can. J. of Psychol. 26*: 140-154.

Klar, Y., & Giladi, E.E. (1997). No one in my group can be below the group's average: A robust positivity bias in favor of anonymous peers. *Journal of Personality and Social Psychology, 73,* 885-901.

Klar, Y., & Giladi, E.E. (1999). Are most people happier than their peers, or are they just happy? *Personality and Social Psychology Bulletin, 25,* 585-594.

Klayman, J., Soll, J. B., González-Vallejo, C., Barlas, S., (1999). Overconfidence: It depends on how, what, and whom you ask. *Organ. Behav. Hum. Decis. Process. 79*, 216-247.

Kober, H., Barrett, L.F., Joseph, J., Bliss-Moreau, E., Lindquist, K., & Wager, T.D.

(2008). Functional grouping and cortical-subcortical interactions in emotion: A meta-analysis of neuroimaging studies. *NeuroImage, 42,* 998-1031.

Koole, S. L., Dijksterhuis, A., van Knippenberg, A., (2001). What's in a name: Implicit self- esteem and the automatic self. *J. Pers. Soc. Psychol. 80*, 669-685.

Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for overconfidence. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 107–118.

Kosslyn, S. M. (1999). If neuroimaging is the answer, what is the question? *Philosophical Transactions of the Royal Society of London: Series B, Biological Sciences, 354,* 1283-1294.

Kriegeskorte N, Simmons WK, Bellgowan PSF, Baker CI. (2009). Circular analysis in systems neuroscience: the dangers of double dipping. *Nat. Rev. Neurosci. 12*: 535-540.

Krienen, F.M., Tu, P.C., Buckner, R.L. (2010). Clan mentality: Evidence that the medial prefrontal cortex responds to close others. *Journal of Neuroscience, 30,* 13906-13915.

Kringelbach, M.L., O'Doherty, J., Rolls, E.T., & Andrews, C. (2003). Activation of the human orbitofrontal cortex to a liquid food stimulus is correlated with its subjective pleasantness. *Cerebral Cortex, 13,* 1064-1071.

Kringelbach, M.L., & Rolls, E.T. (2004). The functional neuroanatomy of the human orbitofrontal cortex: Evidence from neuroimaging and neuropsychology. *Progress in Neurobiology, 72,* 341-372.

Kross, E., Egner, T., Ochsner, K., Hirsch, J., & Downey, G. (2007). Neural dynamics of rejection sensitivity. *Journal of Cognitive Neuroscience, 19*, 945-956

Kruger, J. (1999). Lake Wobegon be gone! The "below-average effect" and the egocentric nature of comparative ability judgments. *J. Pers. Soc. Psychol. 77*, 221-232.

Kruger, J., & Burrus, J. (2004). Egocentrism and focalism in unrealistic optimism (and pessimism). *Journal of Experimental Social Psychology, 40(3*), 332-340.

Krusemark, E.A., Campbell, K.W., Clementz, B.A. (2008). Attributions, deception, and event related potentials: An investigation of the self-serving bias. *Psychophysiology 45,* 511- 515.

Kumashiro, M., & Sedikides, C. (2005). Taking on board liability-focused feedback: Close positive relationshipos as a self-bolstering resources. *Psychological Science, 16,* 732-739.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin, 108,* 480-498.

Kunda, Z., & Sanitioso, R. (1989). Motivated changes in the self-concept. *Journal of experimental social psychology, 25*(3), 272-285.

Kwang, T., & Swann, W.B., Jr. (2010). Do people embrace praise even when they feel unworthy? A review of critical tests of self-enhancement versus self-verification. *Personality and Social Psychology Review, 14,* 263-280

Larwood, L. (1978). Swine flu: A field study of self-serving biases. *Journal of Applied Social Psychology, 18,* 283-289.

Leary, M.R., Haupt, A.L., Strausser, K.S., Chokel, J.T. (1998). Calibrating the sociometer: The relationship between social-evaluative appraisals and the state self-esteem. *Journal of Personality and Social Psychology, 74,* 1290-1299.

Leary, M.R., Terry, M.L., Allen, A.B., Tate, E.B. (2009). The concept of ego threat in social and personality psychology: Is ego threat a viable scientific construct? *Personality and Social Psychology Review, 13,* 151-164.

LeDoux, J.E. (2000). Emotion circuits in the brain. *Annual Review of Neuroscience, 23,* 155-184.

Lench, H., Ditto, P. (2008). Automatic optimism: Biased use of base rate information for positive and negative events. *J. Exp. Soc. Psychol. 44*, 631-639.

Lerner, J.S., Tetlock, P.E. (1999). Accounting for the effects of accountability. *J. Pers. Soc. Psychol. 125*, 255-275.

Lhermitte, F. (1986). Human autonomy of the frontal lobes II. Patient behavior in complex and social situations: The "environmental dependency syndrome." *Ann. Neuro., 19*, 335-343.

Loftus, E.F., & Wagenaar, W.A. (1988). Lawyers' predictions of success. *Jurimetrics Journal, 29,* 437-453.

McCabe, C., et al. (2008). Cognitive influences on the affective representation of touch and the sight of touch in the human brain. *Social Cognitive and Affective*

Neuroscience, 3, 97-108.

McClure, S.M., Li, J., Tomlin, D., Cypert, K.S., Montague, L.M., & Montague, P.R. (2004).Neural correlates of behavioral preference for culturally familiar drinks. *Neuron, 44,* 379-387.

McFarlin, D.B., Blascovich, J. (1981). Effects of self-esteem and performance on future affective preferences and cognitive expectations. *Journal of Personality and Soial Psychology, 40,* 521-531.

McKenna, F.P., Myers, L.B. (1997). Illusory self-assessments – Can they be reduced? *Br. J. Psychol. 88*, 39-51.

Macmillan, N.A., Creelman, C.D. (1991). Detection theory: A user's guide. New York: Cambridge.

Macrae, C.N., Moran, J.M., Heatherton, T.F., Banfield, J.F., Kelley, W.M. (2004). Medial prefrontal activity predicts memory for self. *Cerebral Cortex, 14,* 647-654.

Mason, M.F., Norton, M.I., Van Horn, J.D., Wegner, D.M., Grafton, S.T., Macrae, C.N. (2007). Wandering minds: the default network and stimulus-independent thought. *Science, 315,* 393-395.

Mehta, P., & Beer, J.S. (2009). Neural mechanisms of the testosterone-aggression relation: The role of orbitofrontal cortex. *J. Cogn. Neuro., 22*, 2357-2368.

Mehta, P.H., & Josephs, R.A. (2006). Testosterone change after losing predicts the decision to compete again. *Hormones and Behavior, 50,* 684-692.

Metcalfe, J. (1998). Cognitive Optimism: Self-Deception or Memory-Based Processing Heuristics? *Personality and Social Psychology Review, 2*(2), 100-110.

Milad, M.R., & Quirk, G.J. (2002). Neurons in medial prefrontal cortex signal memory for fear extinction. *Nature, 420,* 70-74.

Miller, D.T., & Ross, M. (1975). Self-serving biases in the attribution of causality: Fact or fiction. *Psychological Bulletin. 82,* 213-225.

Miller, D.T. (1976). Ego involvement and attributions for success and failure. *Journal of Personality and Social Psychology, 34,* 901-906.

Mills, J. (1976). A procedure for explaining experiments involving deception. *Personality and Social Psychology Bulletin, 2,* 3-13.

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

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

Mitchell, J.P. (2009). Social psychology as a natural kind. *Trends in Cognitive Science, 13,* 246-251.

Moore, D., & Small, D. A. (2007). Error and Bias in Comparative Judgment: On Being Both Better and Worse than We Think We are. *Journal of Personality and Social Psychology, 92*(6), 972-989.

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.

Mulder, M.J., Wagenmakers, E.J., Ratcliff, R., Boekel, W., Forstmann, B.U. (2012). Bias in the brain: A diffusion model analysis of prior probability and potential payoff. *Journal of Neuroscience, 32,* 2335-2343.

Murray S.L., & Holmes, J.G. (1997). A leap of faith? Positive illusions in romantic relationships. *Personality and Social Psychological Bulletin, 23*, 586-597.

Neff L.A., & Karney B.R. (2002). Judgments of a relationship partner: Specific accuracy but global enhancement. *Journal of Personality. 70:* 1079-1112.

Neff L.A., & Karney B.R. (2005). To know you is to love you: The implications of global adoration and specific accuracy for marital relationships. *J. Pers. Soc. Psych. 88*: 480-497.

Nisbett, R.E., & Ross, L.D. (1980). Human inference: Strategies and shortcomings of social judgment. Englewood Cliffs, NJ: PrenticeHall.

Nitschke, J.B., Sarinopoulos, I., Mackiewicz, K.L., Schaefer, H.S., & Davidson, R.J. (2006). Functional neuroanatomy of aversion and its anticipation. *NeuroImage, 29,* 106-116.

Northoff, G., Bermpohl, F. (2004). Cortical midline structures and the self. *Trends in Cognitive Sciences, 8*, 102-107

Ochsner, K.N., & Gross, J.J. (2005). The cognitive control of emotion. *Trends in Cognitive Sciences, 9*, 242-249.

Ochsner, K. N., & Lieberman, M. D. (2001). The emergence of social cognitive neuroscience. *American Psychologist, 56,* 717-734.

Ochsner, K.N., Beer, J.S., Robertson, E.R., Cooper, J.C., Gabrieli, J.D.E., Kihlstrom, J.F., & D'Esposito, M. (2005). The neural correlates of direct and reflected self-knowledge. *NeuroImage, 28,* 797-814.

Odean, T. (1998). Volume, volatility, price, and profit when all traders are above average. *Journal of Finance, 8,* 1887-1934.

Oksam, J., Kingma, J., & Klasen, H.J. (2000). Clinicians' recognition of 10 different types of distal radial fractures. *Perceptual and Motor Skills, 91,* 917-924.

Oskamp, S. (1965). Overconfidence in case-study judgments. *Journal of Clinical Psychology, 29,* 261-265.

Ongur, D., & Price, J.L. (2000). The organization of networks within the orbital and medial prefrontal cortex of rats, monkeys and humans. *Cerebral Cortex, 10,* 206-219.

Passingham, R.E., Bengtsson, S.L., & Lau, H.C. (2010). Medial frontal cortex: From self-generated action to reflection on one's own performance. *Trends in Cognitive Sciences, 14,* 16-21.

Paulhus, D. L., Harms, P. D., Bruce, M. N., Lysy, D. C. (2003). The over-claiming technique:Measuring self-enhancement independent of ability. *J. Pers. Soc. Psychol. 84*, 890-904.

Paulhus, D. L., & Harms, P. D. (2004). Measuring cognitive ability with the overclaiming technique. *Intelligence, 32*(3), 297-314.

Paulhus, D., Graf, P., Van Selst, M. (1989). Attentional load increases the positivity of self-presentation. *Soc. Cogn. 7*, 389-400.

Paulhus, D. L. (1998). Interpersonal and intrapsychic adaptiveness of trait self-enhancement: A mixed blessing. *Journal of Personality and Social Psychology, 74*(5), 1197-1208.

Pessoa, L., Japee, S., Sturman, D., & Ungerleider, L.G. (2006). Target visibility and visual awareness modulate amygdala responses to fearful faces. *Cerebral Cortex, 16*, 366–375.

Phan, K.L., Wager, T.D., Taylor, S.F., & Liberzon, I. (2002). Functional neuroanatomy of emotion: A meta-analysis of emotion activation studies in PET and fMRI. *NeuroImage, 16,* 331-348.

Phelps, E.A. (2006). Emotion and cognition: Insights from studies of the human amygdala. *Annual Review of Psychology, 57,* 27-53.

Phelps, E.A., Delgado, M.R., Nearing, K.I., & LeDoux, J.E. (2004). Extinction learning in humans: Role of the amygdala and vmPFC. *Neuron, 43,* 897-905.

Phillips, W.J., Hine, D. W., Thorsteinsson, E. B. (2010). Implicit cognition and depression: A meta-analysis. *Clin. Psych. Rev., 30*, 691-709.

Plassmann, H., O'Doherty, J., Shiv, B., Rangel, A. (2008). Marketing actions can modulate neural representations of experienced pleasantness. *Proceedings of the National Academy of Sciences, 105,* 1050-1054.

Ploghaus, A., Tracey, I., Gati, J.S., et al. (1999). Dissociating pain from its anticipation in the human brain. *Science, 284,* 1979-1981.

Poldrack R.A., & Mumford J.A. (2009). Independence of ROI analysis: Where is the voodoo? *SCAN, 4*, 208-213.

Posner, M. I., & DiGirolamo, G. J. (2000). Cognitive neuroscience: Origins and promise. *Psychological Bulletin, 126,* 873-889.

Price, J.L., & Drevets, W.C. (2010). Neurocircuitry of mood disorders. *Neuropsychopharmacology, 35,* 192-216.

Pronin, E., Gilovich, T., & Ross, L. (2004). Objectivity in the eye of the beholder: Divergent perceptions of bias in self versus others. *Psychological Review, 111,* 781-799.

Randall, D. M., Fernandes, M. F. (1991). The social desirability response bias in ethics research, *J. Bus. Eth., 10*, 805-817.

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation, 20,* 873-922.

Robins, R. W., & Beer, J. S. (2001). Positive illusions about the self: short-term benefits and long-term costs. *J Pers Soc Psychol, 80*(2), 340-352.

Rolls, E.T., Hornak, J., Wade, D., McGrath, J. (1994). Emotion-related learning in patients with social and emotional changes associated with frontal lobe damage. *Journal of Neurology, Neurosurgery, and Psychiatry, 57,* 1518-1524

Rosenberg, M. (1979). *Conceiving the Self.* New York: Basic Books.

Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. *Advances in Experimental Social Psychology, 10,* 173-220.

Ross, M., & Sicoly, F. (1979). Egocentric biases in availability and attribution. *Journal of Personality and Social Psychology, 37*(3), 322-336.

Ross, M., McFarland, C., & Fletcher, G.J.O. (1981). The effect of attitude on recall of past histories. *Journal of Personality and Social Psychology, 10,* 627-634.

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.

Rutter, D.R., Quine, L., & Albery, I.P. (1998). Perceptions of risk in motorcyclists: Unrealistic optimism, relative realism, and predictions of behavior. *British Journal of Psychology, 89,* 681-696.

Sanfey, A.G., Rilling, J.K., Aronson, J.A., Nystrom, L., & Cohen, J.D. (2003). The neural basis of economic decision-making in the ultimatum game. *Science, 300,* 1755-1758.

Sanitioso, R.B., Kunda, Z., & Fong, G. (1990). Motivated recruitment of autobiographical memories. *Journal of Personality and Social Psychology, 57*, 229–241.

Schacter, D.L., Addis, D.R., Buckner, R.L. (2007). Remembering the past to imagine the future: The prospective brain. *Nature Reviews Neuroscience, 8,* 657-661.

Scheibe, C. Ullsperger, M., Sommer, W., Heekeren, H.R. (2010). Effects of parametrical trial-by-trial variation in prior probability processing revealed by simultaneous electroencephalogram/ functional magnetic resonance imaging. *Journal of Neuroscience, 30,* 16709-16717.

Schmeichel, B.J., & Demaree, H.A. (2010). Working memory capacity and spontaneous emotion regulation: High capacity predicts self-enhancement in response to negative feedback. *Emotion, 10,* 739-744.

Schneider, D.J. (1969). Tactical self-presentation after success and failure. *Journal of Personality and Social Psychology, 13,* 262-268.

Schoenbaum, G., et al. (2011). Does the orbitofrontal cortex signal value? *Ann. NY Acad. Sci, 1239,* 87-99.

Sedikides, C. Gregg, A. P. (2008). Self-enhancement: Food for thought. *Pers. Psychol. Sci 3*, 102-116.

Sedikides, C., Herbst, K. C., Hardin, D. P., Dardis, G. J., (2002). Accountability as a deterrentto self-enhancement: The search for mechanisms. *J. Pers. Soc. Psychol. 83*, 592-605.

Sedikides, C., & Green, J.D. (2000). On the self-protective nature of inconsistency-negativity management: Using the person memory paradigm to examine self-referent memory. *Journal of Personality and Social Psychology, 79,* 906-922.

Sellitto, M., Ciaramelli, E., di Pellegrino, G. (2011). The neurobiology of intertemporal choice: Insight from imaging and lesion studies. *Rev Neurosci., 22,* 565-574.

Sharot, T., Riccardi, A.M., Raio, C.M., & Phelps, E.A. (2007a). Neural mechanisms mediating optimism bias. *Nature, 450*, 102-105.

Sharot, T., Martorella, E.A., Delgado, M.R., & Phelps, E.A. (2007b). How personal experience modulates the neural circuitry of memories of September 11. *Proceedings of the National Academy of Science, 104,* 389-394.

Sherman, D.K., Nelson, L.D., & Steele, C.M. (2000). Do messages about health risks threaten the self? Increasing the acceptance of threatening health messages via self-affirmation. *Personality and Social Psychology Bulletin, 26,* 1046–1058.

Sherman, D.K, & Cohen, G.L. (2002). Accepting threatening information: Self-affirmation and the reduction of defensive biases. *Current Directions in Psychological Science, 11,* 119-123.

Shrauger, J.S., Lund, A.K. (1975). Self-evaluation and reactions to evaluations from others. *Journal of Personality, 43,* 94-108.

Singer, T., Critchley, H.D., Preuschoff, K. (2009). A common role of insula in feelings, empathy and uncertainty. *Trends in Cognitive Sciences, 13,* 334-340.

Singer, T., Seymour, B., O'Doherty, J.P., Stephan, K.E., Dolan, R.J., Frith, C.D. (2006). Empathic neural responses are modulated by the perceived fairness of others. *Nature, 439,* 466-469.

Small, D.M., Zatorre, R.J., Dagher, A., Evans, A.C., Jones-Gotman, M. (2001). Changes in brain activity related to eating chocolate: From pleasure to aversion. *Brain, 124,* 1720-1733.

Small, D.M., Gregory, M.D., Mak, Y.E., Gitelman, D. Mesulam, M.M., Parrish, T. (2003). Dissociation of neural representation of intensity and affective valuation in human gustation. *Neuron, 39,* 701-711.

Sommer, K.L., Baumeister, R.F. (2002). Self-evaluation, persistence, and performance following implicit rejection: The role of trait self-esteem. *Personality and Social Psychology Bulletin, 28,* 926-938.

Somerville, L.H., Heatherton, T.F., Kelley, W.M. (2006). Anterior cingulate cortex responds differentially to expectancy violation and social rejection. *Nature Neuroscience, 9,* 1007-1008.

Somerville, L.H., Kelley, W.M., Heatherton, T.F. (2010a). Self-esteem modulates medial prefrontal cortical responses to evaluative social feedback. *Cerebral Cortex, 20,* 3005-3013.

Somerville, L.H., Whalen, P.J., & Kelley, W.M. (2010b). Human bed nucleus of the stria terminalis indexes hypervigilant threat monitoring. *Biological Psychiatry, 68,* 416-424.

Spencer, S.J., Josephs, R.A., & Steele, C.M. (1993). Low self-esteem: The uphill struggle for self-integrity. In R.F. Baumeister (Ed.), *Self-esteem: The puzzle of low self-regard* (pp. 21-36). New York, NY: Plenum.

Spencer, S.J., Fein, S., Lomore, C.D. (2001). Maintaining one's self-image vis-à-vis others: The role of self-affirmation in the social evaluation of the self. *Motivation and Emotion, 25,* 41-65.

Spreng, R.N., Mar, R.A., Kim, A.S. (2009). The common neural basis of autobiographical memory, prospection, navigation, theory of mind, and the default mode: A quantitative meta-analysis. *Journal of Cognitive Neuroscience, 21,* 489-510.

Stanovich, K. E., & Cunningham, A. E. (1992). Studying the consequences of literacy within a literate society: the cognitive correlates of print exposure. *Memory & cognition, 20*(1), 51-68.

Steele, C.M., Spencer, S.J., Lynch, M. (1993). Self-image resilience and dissonance: The role of affirmational resources. *Journal of Personality and Social Psychology, 64,* 885-896.

Steele, C.M. (1988). The psychology of self-affirmation: Sustaining the integrity of the self. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 21, pp. 261-302). New York: Academic Press.

Stein, M.B., Simmons, A.N., Feinstein, J.S., Paulus, M.P. (2007). Increased amygdala and insula activation during emotion processing in anxiety-prone subjects. *American Journal of Psychiatry, 164,* 318-327.

Story, A.L., & Dunning, D. (1998). The more rational side of self-serving prototypes: The effects of success and failure feedback. *Journal of Experimental Social Psychology, 34,* 513-529.

Stricker, L.J., Messick, S., & Jackson, D.N. (1969). Evaluating deception in psychological research. *Psychological Bulletin, 71,* 343-351.

Stuss, D.T., & Benson, D.F. (1984). Neuropsychological studies of the frontal lobes. *Psychological Bulletin, 95,* 3-28.

Suls J., Lemos K., Stewart H.L. (2002). Self-esteem, construal, and comparisons with the self, friends, and peers. *J. Pers. Soc. Psych. 82*: 252.

Suls, J. (1999). The importance of the question in motivated cognition and social comparison. *Psychol. Inquiry 10*, 73-75.

Summerfield, C., & Koechlin, E. (2008). A neural representation of prior information during perceptual inference. *Neuron, 59,* 336-347.

Summerfield, C., Koechlin, E. (2010). Economic value biases uncertain perceptual choices in the parietal and prefrontal cortices. *Frontiers in Human Neuroscience, 4,* 208.

Sutton, R.S., & Barto, A.G. (1998). Reinforcement learning: An introduction. Cambridge: MIT.

Swann, W.B. (2011). Self-verification theory. In P. Van Lang, A. Kruglanski, & E.T. Higgins (Eds.). pp. 23-42. *Handbook of Theories of Social Psychology,* Sage: London.

Swann, W.B., & Read, S.J. (1981). Acquiring self-knowledge: The search for feedback that fits. *Journal of Personality and Social Psychology, 41,* 1119-1128.

Swann, W.B, Hixon, J., Stein-Seroussi, A., & Gilbert, D. (1990). The fleeting gleam of praise: Cognitive processes underlying behavioral reactions to self-relevant feedback. *Journal of Personality and Social Psychology,* 59, 17-26.

Szpunar, K.K., Chan, J.C., McDermott, K.B. (2009). Contextual processing in episodic future thought. *Cerebral Cortex, 19,* 1539-1548.

Takahashi, Y.K., Roesch, M.R., Wilson, R.C., Toreson, K., O'Donnell, P., Niv, Y., & Schoenbaum, G. (2011). Expectancy-related changes in firing of dopamine neurons depend on orbitofrontal cortex. *Nature Neuroscience, 14*, 1590-1597.

Tamir, D.I., & Mitchell, J.P. (in press). The default network distinguishes construals of proximal versus distal events. *Journal of Cognitive Neuroscience.*

Tanaka, S.C., Balleine, B.W., & O'Doherty, J.P. (2008). Calculating consequences: Brain systems that encode the causal effects of actions. *Jornal of Neuroscience, 28,* 6750-6755.

Taylor, S.E., Brown, J.D. (1988). Illusion and well-being: a social psychological perspective on mental health. *Psychological Bulletin, 103,* 193-210.

Taylor, S.E., & Lobel, M. Social comparison activity under threat: Downward evaluation and upward contacts. *Psychological Review, 96,* 569-575.

Taylor S.E., Koivumaki J.H. (1976). The perception of self and others: acquaintanceship, affect, and actor-observer differences. *J. Pers. Soc. Psych. 33:* 403.

Tetlock, P.E., Kim, J.I. (1987). Accountability and judgment processes in a personality prediction task. *J. Pers. Soc. Psychol. 52,* 700-709.

Todorov, A., Baron, S. G., & Oosterhof, N. N. (2008). Evaluating face trustworthiness: A model based approach. *Social, Cognitive, and Affective Neuroscience, 3,* 119–127.

Tracy, J.L., Cheng, J.T., Robins, R.W., & Trzesniewski, K.H. (2009). Authentic and hubristic pride: The affective core of self-esteem and narcissism. *Self and Identity, 8,* 96-213.

Trope, Y. (1986). Self-assessment and self-enhancement in achievement motivation. In R.M.Sorrentino & E.T. Higgins (Eds.), *Handbook of motivation and cognition: Foundations of social behavior* (Vol. 1, pp. 350-378). New York: Guilford Press.

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science, 185*(4157), 1124-1131.

Twenge, J.M., Campbell, W.K. (2008). Increases in positive self-views among high school students: Birth cohort changes in anticipated performance, self-satisfaction, and self-competence. *Psychological Science, 19,* 1082-1086.

Twenge, J.M., & Campbell, W.K. (2003). "Isn't it fun to get the respect that we're going to deserve?" Narcissism, social rejection, and aggression. *Personality and Social Psychology Bulletin, 29,* 261-272.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joilot, M. (2002). Automated anatomical labelling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single subject brain. *NeuroImage, 15*, 273–289.

Uddin, L.Q., Iacoboni, M., Lange, C., Keenan, J.P. (2007). The self and social cognition: The role of cortical midline structures and mirror neurons. *Trends in Cognitive Science, 11,* 153-157.

vanDellen, M.R., Campbell, W.K., Hoyle, R.H., Bradfield, E.K. (2011). Compensating, resisting, and breaking: A meta-analytic examination of reactions to self-esteem threat. *Personality and Social Psychology Review, 15,* 51-74.

Van Lange P.A.M., Rusbult C.E. (1995). My Relationship is Better than-and Not as Bad as-Yours is: The Perception of Superiority in Close Relationships. *Pers. and Soc. Psych. Bull. 21*: 32.

Valentin, V.V., Dickinson, A., & O'Doherty, J. (2007). Determining the neural substrates of goal-directed learning in the human brain. *Journal of Neuroscience, 27*, 4019-4026.

Vincent, J.L., Snyder, A.Z., Fox, M.D., Shannon, B.J., Andrews, J.R., Raichle, M.E., Buckner, R.L. (2006). Coherent spontaneous activity identifies hippocampal-parietal memory network. *Journal of Neurophysiology, 96,* 3517-3531.

Visscher, K.M., Miezin, F.M., Kelly, J.E., Buckner, R.L., Donaldson, D.I., McAvoy, M.P., Bhalodia, V.M., & Petersen, S.E. (2003). Mixed block/event-related designs separate transient and sustained activity in fMRI. *NeuroImage, 19,* 1694-1708.

Vohs, K.D., Heatherton, T.F. (2004). Ego threat elicits differential social comparison processes among high and low self-esteem people: Implications for social-evaluative perceptions. *Social Cognition, 22*, 168-191.

Volkow, N. D., Fowler, J. S, Wolf, A. P., Hitzemann, R., Dewey, S., Bendriem, B., Alpert, R., & Hoff, A. (1991). Changes in brain glucose metabolism in cocaine dependence and withdrawal. *American Journal of Psychiatry, 148,* 621-626.

Vul E., Harris C., Winkielman P., Pashler H. (2009). Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition. *Perspect. on Psych. Sci. 4*: 274-290.

Wallis, J.D. (2007). Orbitofrontal cortex and its contribution to decision-making. *Annual Review of Neuroscience, 30,* 31-56.

Weinstein, N. D. (1980). Unrealistic optimism about future life events. *Journal of Personality and Social Psychology, 39*(5), 806-820.

Whalen, P. J. (1998). Fear, vigilance, and ambiguity: Initial neuroimaging studies of the human amygdala. *Current Directions in Psychological Science, 7,* 177–188.

Wickens, T.D. (2002). Elementary signal detection theory. New York: Oxford.

Wills, T.A. (1981). Downward comparison principles in social psychology. *Psychological Bulletin, 90,* 245-271.

Wood, J.V., Giordano-Beech, M., Ducharme, M.J. (1999). Compensating for failure through social comparisons. *Personality and Social Psychology Bulletin, 25,* 1370-1386.

Zuckerman, M. (1979). Attribution of success and failure revisited, or: The motivational bias is alive and well in attribution theory. *Journal of Personality, 47*(2), 245-287.

Zysset, S., Huber, O., Ferstl, E., von Cramon, D.Y. (2002). The anterior frontomedian cortex and evaluative judgment: An fMRI study. *NeuroImage, 15,* 983-991.

Vita

Brent Laurence Hughes was born in Vancouver, Canada, on January 17, 1981, the son of Laurence Thomas Hughes and Maria Elena Hughes, and the older brother of Natalie Hughes and Jessica Hughes. After graduating from Asociacion Escuelas Lincoln in Buenos Aires, Argentina in 1999, he attended the University of Michigan in Ann Arbor, MI from 1999 to 2003. He graduated from the University of Michigan with a Bachelor of Science in Psychology in December of 2003. Upon graduation, he worked as a lab manager at the Department of Psychiatry at the University of Michigan and co-taught a course in the Department of Psychology at the University of Michigan. In June of 2005, he began a new position as a lab manager at the Department of Psychology at Columbia University in New York City. In August of 2007, he began graduate school, pursuing a doctoral degree in social and personality psychology at the University of Texas at Austin. Upon graduation, he will begin a post-doctoral position at Stanford University.

Permanent email: brencho@gmail.com

This dissertation was typed by the author.