Thesis Presented for the Degree of

DOCTOR OF PHILOSOPHY

in the Faculty of Humanities

Guilt in the Body and Brain:
A Psychophysiological and Neuroimaging Investigation

Melike Marethe Fourie

Department of Psychology

UNIVERSITY OF CAPE TOWN

October 2011

"*It is the confession, not the priest, that gives us absolution*"

– Oscar Wilde (1854-1900)

Declaration

I hereby declare that this submission is my own work, both in concept and execution, and that to the best of my knowledge and belief it contains no material written by another person nor material that has been accepted for the award of any other degree or diploma of the university or other institute of higher learning, except where due acknowledgment has been made in the text.

This thesis contains material published previously, however, the research on which the paper was based is entirely my own, and as first author, I also wrote 95% of the paper.

Melike M. Fourie (Ms)

Date

Acknowledgments

I firstly want to express sincere thanks to my supervisor, Dr Kevin Thomas, who has been a tremendous source of support and guidance throughout my degree, and who has encouraged me to perform independent and novel research. I cannot thank him enough for all his excellent advice and hands-on approach to supervision, he certainly is one in a million.

I am also deeply indebted to my co-supervisors, Prof George Ellis and Associate Prof Ernesta Meintjes, who have encouraged the direction I have taken with this research from the start. In particular, I'd like to thank Prof Ellis for his whole-hearted support and contribution to this research, both conceptually and financially, and for always being accessible to me. I'd also like to thank Prof Meintjes for her invaluable help, especially in terms of the functional imaging data analyses.

I am most grateful to Dr Laurie Rauch for his friendly advice and assistance with physiological data recording and interpretation. In particular, he kindly performed the autoregressive analysis to calculate power spectral densities of heart rate variability for my physiological data analyses. I would also like to thank Dr Barak Morgan for various useful discussions on emotion and behavioural motivation throughout my degree.

In terms of my research topic, I would like to express sincere thanks to Prof Mark Solms, who initially set me on my journey of exploring guilt five years ago. He has also been a major source of inspiration to me, both personally and professionally.

I would also like to thank all the nameless individuals who participated in my research and experienced guilt at the "right time", without their cooperation this thesis would not have been possible.

Finally, I want to thank my loving husband, Rudolph, and family, who have been an immeasurable source of selfless support throughout my studies, and who have waited patiently for me to finish this thesis so that "life as usual" can resume. Without your love, support, and encouragement this work would not have been possible.

This research was supported by the Oppenheimer Memorial Trust, the National Research Foundation, the AW Mellon Foundation, and the University of Cape Town.

Abbreviations

ACC	-	anterior cingulate cortex
ANS	-	autonomic nervous system
BDI-II	-	Beck Depression Inventory II
BA	-	Brodmann area
BIS/BAS	-	Behavioural Inhibition and Behavioural Activation Systems
CMS	-	cortical midline structures
EM	-	emotion manipulation period
fMRI	-	functional magnetic resonance imaging
HF	-	high frequency
HR	-	heart rate
HRV	-	heart rate variability
IAT	-	Implicit Association Test
IMS/EMS	-	Internal and External Motivation to Respond without Prejudice Scales
LF	-	low frequency
mPFC	-	medial prefrontal cortex
OFC	-	orbitofrontal cortex
pACC	-	pregenual anterior cingulate cortex
PANAS	-	Positive and Negative Affective Schedule
PEP	-	preejection period
Post-EM	-	post-emotion manipulation period
pSTS	-	posterior superior temporal sulcus
SCC	-	subgenual cingulate cortex
ToM	-	theory of mind
TOSCA	-	Test of Self-Conscious Affect
TPJ	-	temporo-parietal junction
RMSSD	-	root mean of the squared successive differences
RSA	-	respiratory sinus arrhythmia
SCL	-	skin conductance level
STS	-	superior temporal sulcus
supraACC	-	supragenual anterior cingulate cortex

Table of Contents

List of Figures

List of Tables

ABSTRACT

Guilt in the Body and Brain: A Psychophysiological and Neuroimaging Investigation

Melike M. Fourie

Guilt has been described as a quintessential moral emotion with an important regulatory function for the individual and society. Few studies have, however, empirically investigated guilt, largely because of challenges associated with its real-time elicitation. I aimed to elucidate the physiological and neural correlates of guilt by developing two novel emotion elicitation paradigms. In Study 1, I examined the cardiovascular physiology of guilt and pride in 49 females to uncover physiological substrates underpinning these emotions' behavioural motivations. Although both emotions motivate prosocial behaviour, guilt typically inhibits ongoing behaviour, whereas pride reinforces current behaviour. To succeed in eliciting real emotions, I employed a novel social psychology interaction task. Results pointed to dissociable sympathetic activation during guilt and pride: Guilt participants experienced prolonged cardiac sympathetic arousal as measured by preejection period, whereas Pride participants experienced transient non-cardiac somatic arousal and a shift to low frequency power in the cardiac spectrogram. These findings also supported their distinctive motivational functions. For Study 2, I developed an fMRI social prejudice paradigm using a modified version of the Implicit Association Test. The paradigm employed preprogrammed feedback, indicating either high or non-existent implicit prejudice, to elicit guilt and pride in 22 low-prejudice individuals. fMRI results indicated that the current experience of guilt was associated with increased activity in anterior paralimbic structures (anterior cingulate cortex and anterior insula), as well as in areas associated with mentalising (medial frontal cortex, posterior cingulate, and precuneus). No significant activations were observed for the pride condition. Significant negative correlations between self-reports of guilt and neural activity were also observed in the pregenual anterior cingulate cortex and right posterior insula. These findings may reflect either decreased vagal efference during guilt, or individual differences in a regulatory response to block affective pain. Both studies confirmed that guilt may be reliably elicited with real-time laboratory paradigms. In addition, acute guilt's prominent association with higher self-reported Behavioural Inhibition System sensitivity and conflict-related supragenual anterior cingulate cortex activity provide support for its function as a punishment cue that interrupts ongoing behaviour before ameliorative actions are initiated.

INTRODUCTION

The aim of this thesis was twofold: First, I wanted to demonstrate that guilt may be elicited reliably and measured in real-time through carefully developed emotion elicitation paradigms. Such an approach may be particularly informative because the data obtained would be more ecologically valid than those from, for instance, questionnaire measures typically employed. Second, because guilt is such an important moral emotion, I wanted to characterize its biological substrates to inform our current understanding of the motivational function of this emotion. The studies in this thesis were therefore designed to systematically explore the construct of guilt, at both physiological and neural levels.

Neuroscientists have not always considered the inner landscape of feeling states and consciousness as suitable topics for rigorous empirical research; such subjective matters were typically placed in the realm of psychoanalysis and related disciplines (Solms & Turnbull, 2002). In recent years, however, this situation has changed dramatically, with interest in emotion and affective phenomena assuming center stage in many leading neuroscientific laboratories around the world. Despite this growing interest, the complexities of conducting research in emotion are considerable, most often requiring highly specialized designs and instrumentation along with complex data analytic and methodological challenges, including effective emotion induction procedures (Coan & Allen, 2007). Emotion researchers striving for a complete understanding of any emotional process are therefore faced with significant challenges.

By far the majority of studies in affective neuroscience to date have investigated basic emotion phenomena, perhaps because such emotions are easier to elicit and detect (by way of facial expression), and can also be studied in animals. In the last decade, however, the spotlight has turned to the moral emotions to refine our understanding of the cognitive and neural mechanisms that underlie moral behaviour (Greene & Haidt, 2002; Moll, De Oliveira-Souza, & Zahn, 2008). In this regard, converging evidence indicates that moral intuitions (including moral emotions) are instant feelings that directly influence moral judgments and behaviour (Damasio, 1994; Haidt, 2001). Moral judgment is therefore the result of quick and automatic affective processes, rather than of slow and effortful moral reasoning.

Guilt has been described as a quintessential moral emotion. It is associated with the belief that you have harmed another person and should have thought, felt, or acted differently (Kubany & Watson, 2003). It therefore serves as a powerful motivational force that helps us to distinguish between right and wrong (Kroll & Egan, 2004). Several pitfalls have, however, halted its empirical investigation. Most notably, the problem of eliciting and measuring authentic guilt in the laboratory presents a real challenge. To date, no consensus exists in the literature as to the factors necessary or sufficient to provoke guilt (Baumeister, Stillwell, & Heatherton, 1994). It is a complex social emotion that naturally arises in complex social interactions; situations that provoke guilt may therefore also elicit other negative emotions. Deception may furthermore be necessary in experimental designs of guilt elicitation to evoke an intense emotional experience, which in turn raises ethical concerns (Harmon-Jones, Amodio, & Zinner, 2007).

Despite these difficulties, the empirical investigation of guilt merits investigation because of its prevalence in healthy people and its association with various neurological and psychiatric disorders. Impaired processing of guilt may, for example, lead to amoral or socially inappropriate behaviours, such as those observed in antisocial personality disorder or psychopathy (Blair, 1995; Pardini, Lochman, & Frick, 2003), as well as neuropsychiatric disorders, including frontotemporal dementia and frontal lobe lesions (Krajbich, Adolphs, Tranel, Denburg, & Camerer, 2009; Sturm, Ascher, Miller, & Levenson, 2008). Excessive guilt, on the other hand, has been associated with depressive episodes (O'Connor, Berry, Weiss, & Gilbert, 2002), anxiety disorders (Henning & Frueh, 1997), and obsessive-compulsive disorder (Lewis, 1971; Rachman, 1993).

Empirically disentangling the biological substrates of important moral emotions, e.g., guilt, both physiologically and neurally, could pave the way toward elucidating the mechanisms that underlie moral behaviour, and will advance our understanding of the impact these emotions have on our physical and mental well-being. Such an integrative approach may also foster the synthesis of experimental findings across disciplines. Because autonomic nervous system (ANS) activity is an intrinsic aspect of all emotional responses, the first study in this thesis was designed to elicit and measure physiological responses associated with guilt, with pride serving as a contrasting moral emotion. The second study employed functional neuroimaging (fMRI) to investigate the biological correlates of guilt and pride at a neural level. Both paradigms were designed with the aim of improving on previous investigations, and, in particular, to elicit current emotions relevant to the individual rather than drawing on hypothetical or remembered experiences that may have little ecological validity.

LITERATURE REVIEW

Affective neuroscience is currently one of the most fast-developing and ground-breaking areas of scientific enquiry. Yet, the phenomenon at the heart of this field of study, and implicated in nearly all psychological phenomena, including cognition, memory, perception, and decision-making, remains strangely ill-defined: What exactly constitutes an emotion (Gross, 2010; Lang, 2010)? Across diverse theories, there is some common agreement that emotions may be defined as complex phenomena that consist of coordinated but dissociable subjective, cognitive, physiological, and expressive components that propel us into action and guide goal-directed behaviour.

Within the affective world, however, emotion phenomena may range from reflex-like impulses that occur under the cognitive radar (e.g., an unconditioned fear response), to more complex affective states that require higher-order processing of the environment, self, others, and social standards (e.g., embarrassment). In addition, the time course of affective phenomena may range from momentary emotional reactions, to moods, which last for hours or days, to emotional traits, and, finally to emotional disorders (Ekman, 1994). It is therefore important to make explicit the working definition of emotion that will be employed for the purpose of this thesis. Toward this aim, I define emotions as discrete, transient biological reactions that deal with specific situations in an adaptive fashion by triggering integrated response tendencies or complex behaviours toward the attainment of either survival and reproductive goals, or social goals more indirectly related to survival (Cosmides & Tooby, 2000; Izard & Ackerman, 2000; Tracy & Robins, 2007a). To the extent that emotions direct or alter behaviour, one may expect our affective responses to negative events to be particularly well-developed (Frijda, 1986).

Our affect system has been moulded over millennia of evolutionary forces to serve unique adaptive functions. Emotions trigger survival responses (e.g., fight and flight), environmental exploration, and reproductive behaviours, to name but a few, while a subset of emotional responses have also developed into salient social cues that can betray the individual's current emotional state (Darwin, 1872). Emotional responses are too numerous and nuanced, however, for each to exist as a unique and discrete wiring of the nervous system. Instead, there appears to be a limited set of basic emotion systems that are evolutionarily conserved and that rely on innately specified mechanisms to help us deal with a few species-constant problems related to survival (Levenson, 1999). There is little consensus among emotion researchers, however, on the criteria that should be employed to distinguish these more evolutionarily hard-wired emotions from those that emerge through subsequent learning and cultural experience (Ekman, 1992; Izard, 1992; Panksepp, 1998; Turner & Ortony, 1992). Because the basic or primary emotions are not the main focus of this thesis, I will not comment here further on issues regarding their classification and distinguishing features. For all practical purposes, however, I will employ Ekman's framework whenever I do refer to them. Accordingly, there are six basic emotions: happiness, sadness, anger, disgust, surprise, and fear, which trigger predictable and automatic responses to prototypical environmental challenges, and may be recognized cross-culturally from nonverbal expressions (Ekman, 2003; Ekman & Friesen, 1971; Ekman, Levenson, & Friesen, 1983).

Compared to the basic emotions, the moral and self-conscious emotions are more cognition-dependent and emerge later in ontogeny (Lewis, 2000). They are thought to be unique to humans and certain primates (Gallup, 1982), and play a fundamental role in communal contexts. The formation of social systems is, in fact, considered a key driving force behind the evolution of our current moral apparatus. Evolution favours group-living rather than going solo, because it offers several reproductive and long-term survival advantages, e.g., sharing child-rearing duties and defending against predators (De Waal, 2006). These social functions, in turn, required the evolution of more sophisticated cognitive and emotional capacities, for example, prosocial motives, theory of mind, empathy, and moral emotions like shame and guilt, to enable us to live harmoniously in social groups (Boyd & Richerson, 2009). In neurological terms, an increase in connectivity between affective and cognitive brain structures may have facilitated the evolution of more complex emotional states.

Experiencing affect may lead to many potentially far-reaching consequences, influencing complex phenomena such as moral behaviour, interpersonal relations, and social currency, but also affecting health status (Leary, 2007; Newton, 2009; Tugade, Fredrickson, & Barrett, 2004). In this regard, emotions may be viewed as part of a phenomenal chain, starting with fast-acting central nervous system changes, moving to somewhat slower autonomic nervous system changes, to yet slower neurochemical and hormonal changes, and eventually to changes in health (Larsen & Prizmic, 2004). Although emotions form part of everyday life, frequent exposure to certain emotional states may thus affect short- and long-term physical health and increase risk for psychopathology (Kring & Bachorowski, 1999; Steptoe & Brydon, 2009). To elucidate this well-recognised link between affect and health, ANS and endocrine reactions in responses to discrete affective states have been studied extensively (Kreibig, Wilhelm, Roth, & Gross, 2007; Lerner, Dahl, Hariri, & Taylor, 2007; Rainville, Bechara, Naqvi, & Damasio, 2006; Rottenberg, Kasch, Gross, & Gotlib, 2002).

Psychophysiological research in affective neuroscience has, however, largely neglected the moral emotions, despite their powerful influences on our reasoning, behaviour, and social status. We are ultra-social organisms that stand or fall by our social reputation. Moreover, the moral emotions are arguably more tied to our mental and physical well-being in the complex social environments we have to navigate now in the 21st century, than they have been ever before (see, e.g., Keltner & Haidt, 1999). Consider, for example, the nauseating feeling associated with social exclusion or rejection, the pangs of guilt experienced after wrong-doing, the sheer embodiment of empathic feelings toward another, or the sense of elatedness after a praise-worthy accomplishment. Because these complex emotions form the currency of our social lives, they warrant empirical scientific exploration, despite the fact that the challenges associated with such an endeavour may be enormous.

Guilt as a Moral and Self-conscious Emotion

> *"…morality would be reduced to a meaningless concept if it were stripped from its motivational and emotional aspects"* (Moll, Zahn, de Oliveira-Souza, Krueger, & Grafman, 2005, p. 806).

How do we decide what the right thing is to do in a given situation? This is a prototypical question at the heart of moral psychology, where morality may be defined as a set of rules embraced by a cultural group to regulate interpersonal conduct and cultivate social harmony (Hogan, 1973). The mechanisms that underlie moral behaviour may be parsed into three different domains of enquiry: the moral emotions, theory of mind, and abstract moral reasoning (Casebeer, 2003). In Funk and Gazzaniga's (2009) neurobiological model of human morality, automatic social processing, which includes moral emotions, is thought to activate parallel neural circuits that evaluate the actions and intentions of others and that ultimately results in a decision or action through mechanisms of competition. This process is experienced subjectively as a moral sense of right or wrong. According to their model, moral reasoning is reduced to an interpretive affair that acts post-hoc in an attempt to explain the relation between the subjective moral response and explicit contextual information available. Similarly, most contemporary researchers view affective processes as powerful driving forces in effective moral cognition (Casebeer, 2003; Haidt, 2007; Moll, De Oliveira-Souza, & Zahn, 2008; Nichols, 2002), but also in equitable decision-making (Hsu, Anen, & Quartz, 2008), and the development of appropriate social behaviour (Beer, Heerey, Keltner, Scabini, & Knight, 2003).

The notion that affect is central to moral judgment is an idea that was perhaps re-instigated by the affective revolution of the 1990s, when various researchers gave prominence to automatic, effortless affective mechanisms in terms of their ability to influence cognitive processes (Bargh & Chartrand, 1999; Devine, 1989). Prominent examples include Damasio and colleagues' somatic marker hypothesis (Bechara, Damasio, & Damasio, 2000; Damasio, 1994), and Haidt's (2001) social intuitionist model. Haidt (2001) argued that moral judgment is driven by instantaneous, affect-laden intuitions (e.g., loyalty and reciprocity), and that our actions in the moment are hardly ever influenced by abstract moral reasoning. These convictions, however, were not always held throughout history. Rather, philosophers, for the most part, compared the conflict between reason and emotion to a conflict between divinity and animality (see, e.g., Plato's Timaeus, 4th century B.C./1949). Rationalists since the ancient Greeks have worshipped reason, fostering the classic assumption that higher forms of human existence, including rational thought, may be hijacked by emotion (Cacioppo & Gardner, 1999). In the 18th century, however, Hume's (1777/1960; 1739/1984) attack on rationalism, followed later by the writings of Freud (1901/1965), placed emphasis on intuitive moral sentiments, which they believed are imperative in driving moral judgments and behaviours. The cognitive revolution of the 1960s, nevertheless, had the effect of reinstating purely cognitive models of moral behaviour, i.e., that moral judgment is reached primarily through a process of conscious reasoning (Kohlberg, 1969; Piaget, 1965).

Although people undeniably engage in moral reasoning that can correct and override automatic affect-laden intuitions (Greene & Haidt, 2002; Haidt, 2007), moral psychologists today consider moral emotions to be critical motivational forces that serve to compel immoral behaviour and promote appeasement and reparative actions (Eisenberg, 2000; Keltner & Buswell, 1997). Failure in the appropriate elicitation of these emotions may thus lead to complex social disorders. Various mental (e.g., psychopathy) as well as neurological disorders (e.g., prefrontal damage, frontotemporal dementia), for example, are characterized by a propensity for unacceptable or inappropriate social behaviour (Blair, 2010; Muller et al., 2003; Pardini et al., 2003; Saver & Damasio, 1991; Sturm, Rosen, Allison, Miller, & Levenson, 2006). Impaired moral emotional processing has been proposed to underlie the often observed dissociation in these disorders between *knowing* the right thing to do, and actually *doing* the right thing in real-life situations (Moll, de Oliveira-Souza, Eslinger et al., 2002).

Despite the apparent prominence of moral emotions in guiding appropriate social conduct, very little empirical research has been performed to characterize these emotions. Reasons for this lack of research may be both theoretical and methodological. Moral emotions, for example, show more cross-cultural variability, cannot be distinguished purely based on facial expressions, and are complex affective states that are difficult to capture in artificial experimental situations (Lewis, 2000; Tangney & Dearing, 2002; Tracy & Robins, 2004a). Given the renewed interest in affect in terms of moral behaviour, however, it is now appropriate to develop more accurate functional models and taxonomies of the moral emotions, rather than focusing exclusively on the more "basic" ones. But how should the moral emotions be defined?

Tangney and colleagues view moral emotions as representing a critical element in the link between internalised moral standards and moral behaviour, such that individual differences in the experience and anticipation of these emotions may mediate people's adherence (or lack thereof) to their moral standards (Tangney, Struewig, & Mashek, 2007). Moral emotions differ from basic emotions in that they have more cognitive complexity (Lewis, 2000), and are linked to the welfare of other individuals or of society as a whole (Haidt, 2003). Moral emotions are thus intrinsically interpersonal, because they go beyond the direct interests of the self. Haidt (2003) argued that the more an emotion may be considered as moral, the more easily it should be triggered by an event or elicitor that does not directly concern the self, and the more it should stimulate prosocial action tendencies. Because guilt satisfies both of these criteria robustly, it is not surprising that it is considered a quintessential moral emotion (Eisenberg, 2000).

Also included under moral emotions are envy, empathy, shame, indignation, regret, pride, gratitude, embarrassment, pity, contempt, awe, and jealousy. Moll and colleagues (2007) have proposed a tentative taxonomy wherein moral emotions may be classified as prosocial (e.g., guilt), empathic (e.g., compassion), or other-critical (e.g., contempt). Depending on the specific context, basic emotions, such as disgust and anger, may also acquire moral overtones and qualify as moral emotions (Haidt, 2003).

A few of the moral emotions also belong to a distinct family of ‘self-conscious’ emotions; these primarily include guilt, shame, embarrassment, and pride (Lewis, 2000). They are called ‘self-conscious’ because the self is the object of evaluation, and because they are evoked by either implicit or explicit self-reflection (Tangney et al., 2007). These emotions provide us with immediate and prominent feedback about our current social and moral acceptability, by way of either punishing misbehaviours or reinforcing socially valued behaviours (Tangney et al., 2007; Tracy & Robins, 2004a). Importantly, self-conscious emotions can exert their motivational influences on actual as well as anticipatory behaviours. While the negative evaluation of self (or the behaviour of self) leads to aversive feelings of guilt, shame, and embarrassment, the positive evaluation of self leads to positive feelings of pride and self-approval. The positive or negative acclaim is thus always applied to the self (Williams & DeSteno, 2008).

Self-conscious emotions are intimately connected with other basic emotions, like anger and sadness, but differ from them in that they require self-awareness, and initiate self-evaluative processes according to a set of cultural standards, rules, or goals (Lewis, 2000; Tracy & Robins, 2004a). Leary (2004) pointed out that self-conscious emotions are founded in social relationship and therefore always require a real or imagined audience of other individuals. Moreover, Leary (2004, 2007) contends that these emotions are much more intimately associated with what *other* people may think of us, than with what we think of ourselves. Consider also, for example, James’s (1884, p. 195) articulation on this matter: “The most important part of my environment is my fellow-man. The consciousness of his attitude towards me is the perception that normally unlocks most of my shames and indignations and fears.” This focus on another’s evaluation of self, even if it is an internalised standard of another, is considered unique to self-conscious emotions, and ostensibly serves to maximise interpersonal acceptance and minimise interpersonal rejection (Leary, Koch, & Hechenbleikner, 2001). An important inference from this reasoning is that the ability to evaluate ourselves through the eyes of others requires the ability to represent the mental states of others, that is, ‘theory of mind’ (ToM) or mind reading (Baron-Cohen, 1995; Frith & Frith, 1999). ToM ability is thus considered necessary to recognize or experience self-conscious emotions (Heerey, Keltner, & Capps, 2003; Leary, 2007).

Because self-awareness underlies our ability to read the minds of others (Humphrey, 1986), only animals who possess the capacity of self-awareness (e.g., certain primates) display emotional reactions that may be interpreted as prototypical self-conscious emotions (De Waal, 2001; Gallup, 1982; Parr, Waller, & Fugate, 2005). Similarly, self-conscious emotions are not observed in human infants who have not yet acquired the ability to think consciously about themselves (Lewis, Sullivan, Stanger, & Weiss, 1989). Compared to most basic emotions, which typically develop within the first 9 months of life, the self-conscious emotions emerge considerably later in ontogeny – around the time when a sense of self distinct from others develop (Lewis, 1995). The most basic form of embarrassment, for example, is only observed around 15 to 24 months (Lewis et al., 1989), while more complex self-conscious emotions, like guilt and pride, only emerge during the third year of life (Kochanska, Gross, Lin, & Nichols, 2002; Lagattuta & Thompson, 2007). Self-conscious emotions may also develop later because children first need to understand the standards, rules, and goals that govern culturally appropriate social behaviour (Sullivan, Bennett, & Lewis, 2003). The development of guilt, in particular, is considered closely related to the development of empathic responding (Hoffman, 1982, 1998), and also requires the individual to be able to distinguish clearly between the self and behaviour (Tangney & Dearing, 2002).

The appearance of self-awareness in the evolutionary lineage that led to modern man resulted in dramatic changes in human thought, emotion, and behaviour (Leary, 2004; Leary & Buttermore, 2003). Unlike selfless animals, we are capable of generating emotion only by reflecting upon ourselves and are permitted to imagine how we are being perceived by others. Because all emotion theories since Darwin (1872) have attributed some adaptive advantage to emotional experiences, the fundamental question about the self-conscious emotions is: What is their adaptive significance? Unlike basic emotions, which are essential for survival, Tracy and Robins (2004a, 2007a) argue that the self-conscious emotions are crucial for the attainment of social goals that are tied more indirectly to survival. These emotions facilitate social interaction and maintain interpersonal relationships by virtue of promoting reparative and appeasement behaviours following misdeeds (Baumeister et al., 1994; Keltner & Buswell, 1997; Keltner & Harker, 1998). For example, one is more likely to forgive someone who violated a social convention when that person displays the self-conscious emotion of embarrassment, than someone who displays indifference or no emotion at all (Beer & Keltner, 2004). In addition, self-conscious emotions are adaptive in that they affectively prohibit socially disruptive acts, and that they compel us to behave in socially valued ways to increase the stability of social hierarchies

(Nichols, 2002; Tracy & Robins, 2004a). The self-regulatory function of these emotions in everyday life is thus considered primary (Beer et al., 2003; Beer & Keltner, 2004; Leary, 2007).

Guilt, shame, and other self-conscious emotions. Research on the moral emotions has predominantly focused on the two negatively valenced self-conscious emotions of shame and guilt (Tangney et al., 2007).[1] These emotions are often used interchangeably, and cannot readily be distinguished by the type of situation that led to their elicitation (Tangney, 1992). They are, however, distinct affective experiences associated with very different effects on the individual (Lindsay-Hartz, 1984; Niedenthal, Tangney, & Gavanski, 1994). Because distinctions between shame and guilt have often been confounded in the past, Tangney and Dearing (2002) stressed the importance of clearly differentiating between them in theoretical formulations.

Although guilt and shame have been distinguished based on various criteria (see, e.g., Fontaine et al., 2006), the distinction proposed by Helen Block Lewis (1971) is arguably regarded as the most prominent in literature. According to Lewis, a key difference between guilt and shame centers on the role of the self: whereas guilt focuses on one's bad behaviour ("specific" self-focus), shame focuses on the bad self ("global" self-focus). Both emotions thus involve negative self-evaluations, but while guilt renounces only a particular behaviour, in shame the entire self is apprehended as globally negative. This shift of emphasis from self to behaviour is thought to give rise to fundamentally different phenomenological experiences (Tangney et al., 2007). It should be clarified, however, that while guilt focuses on a specific behaviour, the self is nevertheless evaluated as the agent that performed the negatively apprehended act. Various researchers (e.g., Kubany & Watson, 2003; Lamb, 1983; Lewis, 1971) have, in fact, identified responsibility for causing harm as a critical determinant of guilt. Guilt, like shame, therefore requires a causal link with the self.

The experience of guilt can be described as a painful feeling of discomfort aroused when one has violated a personally relevant moral or social standard, or when another may have such a perception (Baumeister et al., 1994). This aversive feeling is thought to function as an emotional cue that signals when behaviour is unacceptable and should be interrupted or avoided (Baumeister, Stillwell, & Heatherton, 1995; Monteith, 1993). Guilt is also characterized by feelings of tension, especially when people postpone facing up to their guilt (Izard, 1991), as well

[1]Throughout the rest of this thesis I will refer to guilt as either a moral or self-conscious emotion, depending on the specific aspect emphasized.

as a sense of regret or remorse. Regret forms an integral part of guilt as it is the painful realization that things could have been different, yet it may be distinguished from guilt in that it only involves harm to oneself, not to another (Berndsen, van der Pligt, Doosje, & Manstead, 2004). The experience of guilt, however, stops short of the self-condemnation usually experienced with shame; rather, guilt is associated with a desire to change the situation (Niedenthal et al., 1994). Moreover, guilt is usually accompanied by strong empathic feelings toward the victim, and therefore fosters prosocial behaviours to rectify or alleviate the harm one has caused (Ausubel, 1955; Hoffman, 1998). Guilt is thus viewed as more positive and adaptive, and less devastating, than shame, from both an interpersonal and psychological perspective (Tangney & Dearing, 2002).

By comparison, shame has been described as an acutely painful experience where the entire self is painfully scrutinized and experienced as fundamentally flawed (Eisenberg, 2000). Whereas a guilt-ridden person wants to undo aspects of behaviour, the shamed person wants to undo aspects of the self (Niedenthal et al., 1994). In an attempt to escape these painful feelings and avoid further humiliation, shame is associated with the desire to hide, disappear, or die (Ferguson & Crowley, 1997). Shameful feelings are consequently associated with social withdrawal and are thought to disrupt an individual's ability to form empathic connections with others (Tangney, 1991; Tangney et al., 2007). The motivations or "action tendencies" stimulated by shame are thus very different from that of guilt: Whereas guilt motivates the desire to rectify or amend, shame motivates the desire to hide.

Classic and contemporary theorists have also differentiated between guilt and shame based on their relations to prohibitions and ideals, respectively (Freud, 1930/1961a; 1923/1961b; Piers & Singer, 1953; Teroni & Deonna, 2008). The modern-day interpretation of this account is founded in self-discrepancy theory (Higgins, 1987), where guilt is predicted to arise from lack of congruence between actual behaviour and personal ought standards (i.e., prohibitions), whereas shame is predicted to arise from lack of congruence between one's actual behaviour and what one aspires to be based on societal goals (i.e., ideals). Teroni and Deonna (2008) emphasized, however, that shame will only ensue when the ideal in question is actually undermined, because simply not living up to it may be more associated with self-disappointment. Because prohibitions may be equated with norms (something that is forbidden), and ideals with values (something society deems valuable), it follows that guilt is associated with the breaching of a binding norm, whereas shame is associated with undermining a value (Teroni & Deonna, 2008).

One might ask why a certain behaviour sometimes leads to guilt and sometimes to shame? Based on the above distinction, Teroni and Deonna (2008) suggested the following: If a prohibited act is viewed by an individual as the deliberate undermining of a personal value, shame will probably ensue; but if the act is construed as an unintentional *mistake*, guilt will probably ensue. This is because mistakes are less easily viewed as opposing one's values than as dispositions to act against them freely.

Whereas some have declared the different distinguishing criteria between guilt and shame to be conflicting (e.g., Fontaine et al., 2006), Teroni and Deonna (2008) argued that such an opinion is not warranted. Rather, different criteria speak to different dimensions of guilt and shame that together may give rise to a more coherent understanding. For example, guilt may be directed toward behaviour perceived as violating a norm. By comparison, shame is associated with an undermined value, but because values form an integral part of one's self-concept, one's core identity is easily threatened by shame. Based on the above distinction, it also becomes more apparent why shame has sometimes been described as the less 'moral' emotion of the two (e.g., Smith, Webster, Parrott, & Eyre, 2002). Shame's association with ideals rather than prohibitions makes it more likely to arise from nonmoral issues than guilt, e.g., performance failures, inferiority, or the loss of reputation. Shame, nevertheless, serves an adaptive purpose, despite the fact that it is not likely to foster prosocial behaviours. For example, it reminds the individual of standards of propriety that should be adhered to within a social group (Ferguson & Crowley, 1997). In addition, shame has been associated with a coordinated physiological response in reaction to "social self" threats, which may trigger adaptive behaviours in certain contexts (Dickerson, Gruenewald, & Kemeny, 2004).

To conclude the section on guilt and shame, something should also be said on their relation to psychopathology – a topic of active debate. Guilt, in principle, has been argued to have both adaptive and maladaptive consequences (Kugler & Jones, 1992). Tangney and colleagues' cumulative research, however, seems to suggest that "pure guilt", untarnished by shame, serves purely adaptive purposes for the individual (Tangney & Dearing, 2002; Tangney et al., 2007; Tangney, Wagner, & Gramzow, 1992). In this vein, guilt fosters moral behaviour through reparative and empathic behaviours. By comparison, shame has been associated more consistently with negative or pathogenic consequences, e.g., aggression, irritability, externalization, and substance abuse (Dearing, Stuewig, & Tangney, 2005; Dickerson, Kemeny, Aziz, Kim, & Fahey, 2004; Kim, Talbot, & Cicchetti, 2009; Tangney, Wagner, Fletcher, & Gramzow, 1992). Results on the topic of psychological symptoms and adjustment, however, vary widely (see, e.g., Ferguson, Stegge, Eyre, Vollmer, & Ashbaker, 2000; Harder, 1995), with a possible confounding factor being the use of different indices of guilt and shame. For example, scenario-based measures generally highlight the positive potential of guilt, whereas measures that assess dispositional guilt that is more chronic or ruminative generally find little difference between shame and guilt in terms of emotional and social adjustment (Einstein & Lanning, 1998; Eisenberg, 2000).

Nevertheless, both emotions may eventually become maladaptive when the emotional experiences are inappropriate or excessive (Lewis, 1971; Quiles & Bybee, 1997). Clinicians, for example, often associate excessive guilt and shame with the obsessive mental undoing of past emotional events (Niedenthal et al., 1994). More specifically, guilt has often been described as an agitation-based emotion that is associated with anxiety and obsessive-compulsive disorder (Ferguson & Crowley, 1997; Lewis, 1971; Rachman, 1993). By comparison, shame has been described as a dejection-based emotion that plays an important part in depression (Lewis, 1971; Orth, Berking, & Burkhardt, 2006; Thompson & Berenbaum, 2006). More research employing measures that clearly distinguish between situational and dispositional affect, however, are necessary to tease apart the long-term effects of guilt and shame on mental health.

The last negative emotion in the family of self-conscious emotions is embarrassment. Embarrassment has often been conceptualized as a milder form of shame (Izard, 1977; Lewis, 1971), yet recent data suggest that it may be classified as a distinct emotional response that cannot be distinguished purely by affect intensity (Keltner & Buswell, 1997; Tangney, Miller, Flicker, & Barlow, 1996). Whereas guilt and shame are associated with the violation of moral standards, embarrassment appears to be less central to the domain of morality (Tangney et al., 2007). Rather, embarrassment is associated with the (often accidental) violation of social conventions (e.g., etiquette and hygiene, choices of clothing, faux pas, cognitive errors, and loss of bodily control) (Haidt, 2003). Its function in these situations is thought to appease others and promote adherence to social norms (Keltner, 1995; 2000). Embarrassment is furthermore a uniquely public emotion that requires either the real presence of others, or imagining vividly what others may think of you (Miller, 1996; Tangney et al., 1996).

Finally, the positive self-conscious emotion of pride has long been neglected in psychological research. It arises when people believe they are socially valued or responsible for socially valued acts (Mascolo & Fischer, 1995). Pleasurable feelings of pride are considered to serve essential motivational functions in everyday life: pride rewards and reinforces achievement-oriented as well as socially valued behaviours, thereby stimulating future pride-eliciting behaviours (Tangney et al., 2007). In addition, the motivational function of pride has been demonstrated to act as a mechanism that helps us persevere on effortful tasks to achieve long-term goals, despite short-term costs (Williams & DeSteno, 2008). Taken together, pride fuels our human desires and pursuits, while bestowing social status and acceptance upon those individuals who perform such socially valued acts (Tracy & Robins, 2007a). Another social function of pride may thus be to inform others of our status. Pride is also intrinsically linked to the development and maintenance of a genuine sense of self-esteem, which in turn informs individuals of their social value (Brown & Marshall, 2001; Mascolo & Fischer, 1995).

Tracy and Robins (2004a, 2007b) have distinguished between two distinct forms of pride that are associated with very different psychological correlates. Paralleling Lewis's (1971) self versus behaviour distinction, they identified 'hubris' as global pride in self, whereas 'authentic' pride is more associated with event-specific achievements. Hubris or pridefulness is largely viewed as maladaptive and is associated with negative personality traits, e.g., narcissism, hostility, and shame-proneness. By comparison, the achievement-oriented form of pride is associated with adaptive traits (e.g., conscientiousness and extraversion), and promotes further successes and prosocial behaviours (Tracy & Robins, 2007a).

The Experimental Elicitation of Emotion: Requirements and Caveats

> *"I myself took for granted without discussion that the word 'emotion' meant the rank feeling of excitement, and that the special emotions were names of special feelings of excitement, and not of mild feelings that might remain when the excitement was removed.*" – (James, 1894, p. 208).

The elicitation of authentic and truly felt emotional responses in the laboratory is certainly one of the most challenging aspects of emotion research in general (Coan & Allen, 2007). This statement is particularly true of the self-conscious emotions, because methods typically employed to elicit basic emotions appear less effective in eliciting these more inward sensitivities (Tracy & Robins, 2004a). Although considerable research and attention has been devoted to self-conscious emotions, they are cognitively complex and typically arise in social contexts through intricate, self-evaluative processes (Lewis, 2000; Tracy & Robins, 2008). Self-conscious emotions, therefore, cannot be distinguished purely by the nature of external stimuli, or by the types of situations that elicit them, for it is rather the manner in which individuals perceive the situation that defines the resulting emotion (Lewis, 2000). Moreover, self-conscious emotions may be strongly influenced by the cultural context and personal beliefs of participants (Bierbrauer, 1992; Haidt, 2003). Ecological validity in the investigation of such emotions is thus of primary importance (Casebeer, 2003; Moll, Zahn et al., 2005).

A distinct challenge for most emotion induction methods, and particularly those pertaining to the moral emotions, is to create experimental manipulations that involve more than mere intellectual contemplation of a scenario. For it to qualify as a real emotional experience, distinct from cold reason, participants also have to experience a "bodily reverberation" or "feeling of excitement" of some kind (James, 1884, 1894). Moral emotion elicitation paradigms to date may typically be divided into three categories: (i) those where participants have to imagine what they would do in highly unrealistic/unfamiliar scenarios, which often involve deciding whether they would respond in a utilitarian manner for aggregate welfare or not (e.g., "Would you smother your own baby to save yourself and other townspeople from enemy soldiers?") (Greene, Nystrom, Engell, Darley, & Cohen, 2004; Greene, Sommerville, Nystrom, Darley, & Cohen, 2001; Koenigs et al., 2007; Valdesolo & DeSteno, 2006), (ii) those where participants have to imagine their reactions to more realistic moral scenarios, albeit not related to them personally (e.g., "You drive home from a party one evening and crash into another car at a stop light and kill the passengers") (Berthoz, Armony, Blair, & Dolan, 2002; Finger, Marsh, Kamel, Mitchell, & Blair, 2006; Kédia, Berthoz, Wessa, Hilton, & Martinot, 2008), (iii) and those experimental contexts, by far in the minority, where participants are placed in real, emotion-evoking situations that pertain to themselves (e.g., Herrald & Tomaka, 2002).

Although the first two categories may generate moral emotions with some degree of success, they can hardly be taken as a proxy for the type of emotion that occurs in everyday life: Judging oneself based on fictitious mental representations is likely to yield a different emotional evaluation compared to judgments based on actual personal experiences (Kédia et al., 2008).

Rather, it is likely that experiments in the last category hold the most merit in terms of eliciting ecologically valid emotions. To support this claim, I provide a brief overview of emotion elicitation methods that have been employed in a wide variety of previously-published studies.

Emotion elicitation methods in previous research have included deliberate mood-induction, e.g., “Think of things that make you sad” (George, Ketter, Parekh, Herscovitch, & Post, 1996); presentation of emotionally evocative stimuli, such as the International Affective Picture system (IAPS) (Bradley, Codispoti, Cuthbert, & Lang, 2001; Lane, Fink, Chau, & Dolan, 1997), viewing emotive facial expressions (Ekman & Friesen, 1976; Vuilleumier & Pourtois, 2007), or emotional film segments (Britton, Taylor, Berridge, Mikels, & Liberzon, 2006; Kreibig et al., 2007), reading short emotion sentences (Moll et al., 2007; Takahashi et al., 2004) or vignettes (Berthoz et al., 2002); listening to emotional music segments (Khalfa, Isabelle, Jean-Pierre, & Manon, 2002); contorting faces into prototypical emotion expressions (Coan & Allen, 2003b; Ekman et al., 1983); and recording previous emotive experiences (script-driven imagery), where participants are required to recall and relive a past emotive experience while listening to an audio-taped description of the event (Dougherty et al., 1999; Shin et al., 2000).

Although each of these methods has its own merits, most of them may have limited ability to elicit a genuine or intense emotional experience capable of arousing the autonomic nervous system (Herrald & Tomaka, 2002; Stemmler, Heldman, Pauls, & Scherer, 2001). Reading short emotive sentences or vignettes, for example, may be associated with significant subjective reports of emotion but limited activation of the autonomic nervous system, because participants may be making emotional judgments in line with the emotional characteristics of the task, rather than experiencing real affect (Levenson, 2003a). Research furthermore suggests that people often overestimate the intensity of their emotional reactions to hypothetical scenarios (Wilson & Gilbert, 2005). In contrast, Ekman and colleagues’ Directed Facial Action Task (DFA) (Ekman et al., 1983; Levenson, Ekman, & Friesen, 1990) has limited capacity to induce strong phenomenological experiences of emotion in participants (Boiten, 1996), although it does appear to elicit significant emotion-specific autonomic activity (Levenson & Ekman, 2002). Various complex affective phenomena, including guilt, are not accompanied by clearly distinguishable facial expressions, however (Robins, Noftle, & Tracy, 2007), rendering the DFA unsuitable to study such phenomena.

Emotion induction through use of emotion-laden films is a particularly popular approach used to study emotion in a well-controlled laboratory environment (see, e.g., Hubert & de Jong-Meyer, 1991; Kreibig et al., 2007). This method offers a reliable, replicable, and ethically acceptable way to elicit strong emotional reactions across several response systems (including experience, behaviour, and physiology) when standardised film clips are used (Rottenberg, Ray, & Gross, 2007). Despite these practical advantages, however, several factors also limit the usefulness of film stimuli in emotion elicitation. Notably, films are high in cognitive demand and attentional capture, which are associated with well-recognized autonomic correlates (e.g., heart rate deceleration) that will impact on the overall autonomic response (Levenson, 2003a). The ecological validity of film stimuli has also been questioned. Frijda (1989), for example, argued that emotions elicited by imaginary events are unlike emotions elicited by comparable real-life events, and require a willing "suspension of disbelief" of the illusory reality of the film for its operation. Film-induced emotions are furthermore experienced from a characteristic observational angle, i.e., a knowing that the self is not directly concerned (Frijda, 1989). In this regard, film-induced emotions may be similar to emotions evoked in real-life observational situations.

When it comes to ecological validity, especially for the more social emotions, imagery or reliving a previous emotional experience may seem like a particularly fine approach, because the experience remembered is personally relevant and constitutes an event that each individual finds personally arousing. Waldstein et al. (2000), for example, found that cardiovascular arousal for both happiness and anger was more pronounced during a personally-relevant emotion recall condition than during a film-viewing condition. Several emotion researchers, however, have pointed out the failure of imagery paradigms to elicit reliable emotion-specific physiological differences, or to produce similar ANS patterns as other methods (Cacioppo, Berntson, Larson, Poehlmann, & Ito, 2000; Stemmler, 1989; Zajonc & McIntosh, 1992). Such findings may be explained in several ways. Most notably, re-experiencing an emotion in a controlled and safe experimental environment is clearly phenomenologically different from the original emotional encounter. For example, in reliving a previous fearful experience (e.g., a bank robbery), one may experience only a memory or diminished version of the emotion, or experience nothing at all (Herrald & Tomaka, 2002). In line with this view, the intensity of relived emotions often appears to be too low to detect somatovisceral activation (see, e.g., Shin et al., 2000; Stemmler et al., 2001), and may also be affected by an individual's ability to relive a previous episode.

Furthermore, it is unclear what the effect of reliving different personal experiences may have on experimental results (Rainville et al., 2006). Memories, for example, are often emotionally mixed, such that each participant's experience may have a different emotional overlay, e.g., sadness coupled with guilt, versus sadness coupled with anger. Such differences in the experienced emotion may in turn create unwanted variation in the emotions recorded.

In light of the limitations of the methods described above, contemporary emotion researchers stress the importance of moving away from the study of remembered or hypothetical experiences and moving toward the examination of emotions in vivo, *during* meaningful personal interactions that have higher internal validity than hypothetical or past events (Herrald & Tomaka, 2002; Smith & Pope, 1992; Williams & DeSteno, 2008). Fischer and Kleef (2010), for example, recently pointed out emphatically that emotions are inherently social processes, and that the social dimension should be more incorporated in emotion research. Social emotions, in particular, critically depend on the social environment and need to be evoked in the contexts in which they naturally occur (Harmon-Jones et al., 2007; Roberts, Tsai, & Coan, 2007). This social dependence is moreover true of guilt, which has been described as an "interpersonal phenomenon that is functionally and causally linked to communal relationships between people" (Baumeister et al., 1994, p. 243). The real-time elicitation of interpersonal guilt thus requires a paradigm involving complex social interactions, yet such manipulations have remained largely elusive to date because of various methodological challenges.

The study of emotion has long embraced social psychology experiments that employ well-constructed cover stories to mask deception (Aronson, Ellsworth, Carlsmith, & Gonzales, 1990; Gambaro & Rabin, 1969; Harmon-Jones et al., 2007; Harmon-Jones & Sigelman, 2001; Schachter & Singer, 1962). Because such experimental manipulations are representative of the natural contexts in which human emotions typically occur, they have high ecological validity, with dyadic interaction tasks considered to have the greatest such validity (Roberts et al., 2007). Although social psychology experiments are often difficult to conduct and require much planning and creativity, they can be high in realism and have the potential to elicit valid emotional responses that are unaffected by social desirability concerns. Researchers are, however, often challenged by the lack of experimental control afforded by experiments that resemble real-life situations (Levenson, 2003a). The duration of events in such situations, for example, may be hard to control across experiments, particularly if interpersonal dialogue is involved. In addition, social psychology experiments that induce emotions may pose serious ethical concerns: Not only do they employ deception, but the elicitation of strong negative affect may be experienced as painful or traumatic by research participants (see Aronson et al., 1990, for a discussion of these ethical issues).

Despite the above concerns, deception may be unavoidable if important psychological phenomena that impact on health and society, such as guilt, are to be probed. Harmon-Jones and colleagues (2007) have expressed their opinion on this matter quite strongly: They argued that methods that fail to elicit valid emotions, e.g., hypothetical scenarios, should also be considered ethically questionable, because valuable resources may be spent to obtain meaningless or misleading data.

High-impact real-life manipulations, however, need not cause lasting distress if participants are treated with care and sensitivity. To avoid potential negative repercussions, experimenters should pay special attention to post-experimental interviews. During these interviews, each participant should be debriefed adequately about the true purpose of the experiment and why deception was necessary, and, importantly, the experimenter should ensure that each participant leaves feeling positive about his or her experimental experience. Post-experimental interviews are also vital in determining whether a participant's data should be included in subsequent data analysis (Levenson, 2003a), i.e., to ascertain whether the deception was effective and to verify the nature of the experienced affect.

Emotion elicitation in the scanner. With regard to neuroimaging studies of emotion, researchers are faced with several additional challenges. Notably, the highly artificial nature of the scanner environment places several constraints on the natural elicitation of desired emotional states. Neuroimaging requires participants to be placed in an environment that may be anxiety-inducing in and of itself for many; in the case of fMRI, participants also have to concentrate on not making any head movements during the experiment.

Researchers are also severely constrained by methodological limitations of neuroimaging. For example, imaging studies mostly require several instances of the same emotion to be aggregated in order to create enough power to detect significant brain activation. However, repeated exposure to emotional stimuli may be associated with habituation effects (Liberzon et al., 2000). The scanner environment also places obvious limitations on the types of emotion manipulations possible; because of the highly controlled environment, ecologically valid paradigms are difficult to construct.

Perhaps because of these difficulties and constraints, there has been a tendency in fMRI emotion paradigms to investigate dimensions of emotion, rather than discrete affective states (e.g., Canli et al., 2001; Hamann, Ely, Hoffman, & Kilts, 2002; Lane, Fink et al., 1997; Liberzon, Phan, Decker, & Taylor, 2003; Sabatinelli, Lang, Bradley, Costa, & Keil, 2009; Viinikainen et al., 2010). It is easier to manipulate dimensions of emotion, like arousal and valence, in the scanner (e.g., by using emotive pictures or faces) than it is to evoke discrete emotional states relevant to the participant. Imaging studies on basic as well as moral emotions, however, suggest that brain activations associated with distinct positive and negative emotions cannot be explained solely on the basis of valence (Britton, Phan et al., 2006; Zahn, Moll et al., 2009).

Several additional features of moral emotions make them particularly difficult to study with neuroimaging techniques. Casebeer (2003), for example, pointed out that moral cognition is genuine, in that our moral cognitive equipment, which consists of emotion, reason, and action, evolved to effectively coordinate *actual* behaviour, not hypothetical scenarios. It is difficult, however, to create such believable situations within the scanner, where participants would perceive themselves to be agents of current moral/immoral actions. As mentioned previously, moral emotions are also social phenomena, yet real social interaction is difficult to simulate within the scanner.

Several researchers have tried to overcome the lack of social interaction afforded by the scanning environment by creating novel paradigms where participants interact with either real or imagined others. Prohovnik and colleagues, for example, made use of simulated face-to-face social interactions to induce affect. Participants were shown video clips of actresses describing personal experiences of happiness/sadness while looking directly at the camera (viewer) (Prohovnik, Skudlarski, Fulbright, Gore, & Wexler, 2004). The authors argued that their paradigm offered multifaceted emotional stimulation similar to real-life situations, which would thus induce affect in participants through automatic mimicry. Their results suggested that socially relevant emotional stimulation evokes robust neural responses in participants. A different approach was adopted by Eisenberger and colleagues, who employed an effective fMRI paradigm to study the neural effects of social exclusion (Eisenberger, Lieberman, & Williams, 2003). Their task consisted of a virtual ball-tossing game ("Cyberball") during which participants were deceived into believing that they were playing ball with two fellow players (ostensibly in different scanners). After some time, however, the participant became ostracized (excluded) from the game when the two players stopped throwing him/her the ball. The Cyberball paradigm has also recently been adopted to study empathic responses (Masten, Morelli, & Eisenberger, 2011), as well as jealousy (Harmon-Jones, Peterson, & Harris, 2009).

Other imaging paradigms on moral values have developed economic games to study brain activation during the real-time social interaction of two or more players (de Quervain et al., 2004; Sanfey, Rilling, Aronson, Nystrom, & Cohen, 2003). Even more remarkable is the hyper-scanning methodology that has been employed by Montague et al. (2002), which can be used to study the social interactions of several people while being scanned simultaneously. Creating a social dimension, or the illusion thereof, in neuroimaging paradigms is therefore difficult, though not impossible. Nevertheless, emotion researchers endeavouring to uncover the brain organisation of emotion face serious challenges. Innovative studies that effectively overcome the lack of social interaction during fMRI will no doubt be essential in elucidating the neural correlates of moral emotion.

Taken together, the difficulty with many of the methods described above is that they may not elicit anything resembling an actual emotional experience that one would encounter outside the laboratory (Herrald & Tomaka, 2002). In this regard, the key to true, purposeful, and intense emotion elicitation would be to create an ecologically valid situation in which the participant experiences the relevant emotion as a feeling pertaining to the self (Harmon-Jones et al., 2007). This approach would undoubtedly also assist investigations of emotion-specific autonomic activity, an area of much controversy and debate.

The Measurement of Emotion: Psychophysiology

> ”…*emotions briefly take the reins of the ANS and alter its pattern of activation in service of behaviours that are likely to deal successfully with particular problems/challenges that face the organism*.” (Levenson, 2003a, p. 219)

Another key factor hindering the systematic empirical investigation of self-conscious emotions is their measurement. Guilt is an internal affective state that is difficult, if not impossible, to assess directly (Tangney, 1996). Whereas the other self-conscious emotions (embarrassment, pride, and shame) are associated with recognizable patterns of nonverbal behaviour (Keltner, 1995; Keltner & Buswell, 1997; Tracy & Robins, 2004b, 2008), guilt has been described as the negative emotion with the least distinct facial expression (Izard, 1991). Although guilt may also be associated with changes in outward appearance, emotional responses may occur without overt facial or behavioural signals if the experienced emotion is subtle or inhibited (Ekman, 1993). The assessment of guilt therefore often relies extensively on self-reported feelings. While subjective evaluations have previously been considered the most suitable data for the study of emotion, they only provide an incomplete glimpse of the underlying biological structures and processes associated with emotion (Cacioppo, 2004). This lack of objectivity has hampered progress in the past, placing emotions in the domain of philosophical and theoretical speculation. Researchers are now, however, advocating the importance of using more objective measures to construct and test theories of emotion if we are to expand our knowledge of emotion-related phenomena (Coan & Allen, 2007).

Although the need for more objective measures has been emphasized in several critiques of self-report (see, e.g., Nisbett & Wilson, 1977; Ryaff & Singer, 2003), a large section of emotion research continues to depend entirely on subjective measures to derive and test theories. Problems associated with self-report emotion measures include the fact that they require participants to be both aware of and willing to disclose their emotions, assumptions that are frequently not met if emotions are experienced implicitly, or if participants are unwilling to openly discuss their feelings (Robins et al., 2007). Language may also present a problem in capturing the precise affective experience. Lay-people, for example, frequently confuse the distinction between similar emotions like guilt and shame (Tangney & Dearing, 2002). Emotion self-reports are also vulnerable to demand characteristics and can easily be distorted to appear more socially desirable (Harmon-Jones et al., 2007). Males and females may also distort emotional responses in different ways to conform to gender stereotypes (Brody & Hall, 2000). Finally, the response coherence across different responses systems (e.g., self-report and physiology) may be limited (Mauss, Levenson, McCarter, Wilhelm, & Gross, 2005). Self-reports can therefore not be used, *ipso facto,* to infer responses in other systems. Though essential to capture some aspects of the phenomenality of emotion (Nielsen & Kaszniak, 2007), self-reports of emotion thus have the potential to bias data in emotion research, and should ideally be used in conjunction with other measures to capture the emotional experience adequately.

In the search for more objective measures to probe and track emotional processes, psychophysiological methods have become part of most mainstream research laboratories. Contemporary emotion researchers increasingly make use of continuous physiological recordings to better understand the underlying psychobiological processes at work during an emotional response, and to inform theories of emotion (Curtin, Lozano, & Allen, 2007). Although cognitive science has taught us much about the ubiquitous nature of emotion in all aspects of human cognition and behaviour, emotions are essentially physiological processes that cannot be fully understood without considering their underlying physical substrates (Cacioppo & Gardner, 1999). Changes in indices of visceral and cardiovascular autonomic activity may thus be interpreted as objective indicators of emotional processing. Moreover, because somatovisceral activation constitutes the interface between brain and behaviour, studying the physiology of emotion may also indirectly inform theories of their psychological function (Stemmler et al., 2001).

Moral emotions have largely been neglected when it comes to their psychophysiological measurement. James, more than a century ago, hinted at the association between the "inward sensibilities" (e.g., indignation and pride) and physiological change, reasoning that those feeling states may at first sight appear devoid of all bodily reverberation, yet are "fruits of the same soil (as the 'standard' emotions) with the grossest bodily sensations of pleasure and of pain" (James, 1884, p. 201). No study to date, however, has investigated the cardiovascular physiology of a current guilt response, and only recently has research suggested that the self-conscious emotions of embarrassment, shame, and pride are associated with distinct psychobiological responses (Dickerson, Gruenewald et al., 2004; Herrald & Tomaka, 2002; Keltner & Buswell, 1997). Moral or self-conscious emotions' association with peripheral physiological responses is hardly surprising, however, when one considers the strong motivational forces that these emotions exert on our daily behaviours (Kroll & Egan, 2004). Moreover, distinct moral emotions have been associated with disease pathogenesis, which suggests that they may trigger bodily changes as early precursors of the disease process (Brosschot & Thayer, 1998; Mayne, 1999; Orth et al., 2006; see also review in Tangney et al., 2007). The psychophysiological exploration of moral emotions thus warrants further investigation.

Autonomic specificity. The existence and role of unique somatovisceral activation patterns for individual emotions is a controversial and methodologically challenging issue in emotion research (Davidson, Jackson, & Kalin, 2000; Ellsworth, 1994; Levenson, 1992; Stemmler, 1989). It may well be described as one of the most enduring topics in psychology, given that it has spurred a large amount of research and debate over the last century.

Although Darwin (1872) noted that bodily reactions to specific emotional stimuli appear to be conserved across species, the debate may have begun properly with James's (1884) assertion that emotional experiences are associated with unique patterns of bodily responses, and that the perception of these peripheral physiological changes *is* the emotional experience. What became known as the James-Lange theory of emotion (James, 1894; Lange, 1885) thus has a slightly backward twist to it: The emotional experience is not one's mental response to a stimulus, but rather the perception of the ensuing peripheral physiological changes that result in an emotional feeling. Accordingly, we do not tremble because we feel afraid, but we feel afraid because we tremble; the emotional colour of any feeling state is due to the brain's perception of bodily arousal. Neural representations of automatically generated bodily responses have also more recently been incorporated in theories that posit a central role of autonomic feedback (expressed as "gut feelings") in guiding motivational behaviour and social interaction (Damasio, 1994, 1999; Katkin, Wiens, & Öhman, 2001).

Probably the most well-known opposition to the James-Lange theory of emotion is that of Walter Cannon (1927, 1929), whose most salient argument was that bodily responses are too undifferentiated and stereotyped to provide the fine distinctions necessary to account for the wealth of emotional feelings. Rather, changes in bodily arousal manifest as generalised, nonspecific sympathetic arousal (e.g., Schachter & Singer, 1962), and are not necessary/critical for the emotional experience. By this account, the emotional experience is thus exclusively mediated by central networks.

Despite considerable evidence outlining the presence of highly differentiated and organ-specific response patterns within both axes of the autonomic nervous system (Morrison, 2001; Porges, 1995), the issue of whether afferent signals from the body are integral to the experience of unique emotional states remains unresolved (Heims, Critchley, Dolan, Mathias, & Cipolotti, 2004). In a recent attempt to refute Cannon's heuristic, Harrison and colleagues showed that organ-specific autonomic responses during two types of disgust differentiated central emotional feeling states (Harrison, Gray, Gianaros, & Critchley, 2010). Changes in both cardiac and gastric activity also correlated with the overall magnitude of experienced disgust. In particular, however, they identified, through use of fMRI conjunction analyses, a common neural substrate (anterior insula) that was responsive to both physiological changes and the subjective experience, thus functioning as an interface between embodied and experiential processes.

Most psychologists today will acknowledge the role of peripheral feedback in modulating emotional intensity; a point of great contention in emotion research, however, is the notion that every emotion is autonomically unique (Critchley, 2005). In this regard, it should be noted that autonomic signatures may be observed only during the brief period after the arousal of an emotion and before actual behaviour commences (Levenson, 2003a; Stemmler, Aue, & Wacker, 2007). Autonomic specificity is therefore not expected to pertain to longer emotional phenomena, e.g., mood states. Furthermore, specificity does not require every emotion to be associated with a unique ANS signature, but only that some affective states differ from others in consistent ways (Levenson, 1992). James himself, for example, did not assert that the physiological responses, or, for that matter, the symptoms of any particular emotion, be set in stone; he argued that bodily variation may occur within limits that still preserve the *functional* resemblance of the given emotion (James, 1894). The search for autonomic specificity in current emotion research may thus have been born out of an unwarranted focus on a single aspect of James's theory (Ellsworth, 1994). Perhaps more informative would be to investigate under what conditions and for which emotions differential physiological activity exists (Cacioppo et al., 2000).

Overall, good evidence exists for distinct physiological patterning of various basic (and primarily negative) emotions like anger, fear, and disgust (Cacioppo et al., 2000; Collet, Vernet-Maury, Delhomme, & Dittmar, 1997; Ekman et al., 1983; Levenson et al., 1990; Rainville et al., 2006). Heart rate appears to discriminate best among these emotions (Zajonc & McIntosh, 1992), with greater heart rate acceleration typically observed for anger, fear, and sadness compared to disgust. By comparison, autonomic specificity for positive emotions remain uncertain, an issue perhaps exacerbated by the fact that they have received far less attention than negative emotions in psychophysiological research, and are often studied individually (Levenson, 2003a).

Emotion specificity research is, however, plagued by findings that do not always replicate across studies, or across different emotion elicitation procedures (Boiten, 1996; Stemmler, 1989). In this regard, variability associated with different emotion elicitation techniques and selected autonomic measures have certainly contributed to some of the inconsistencies that characterize research in this area (Christie & Friedman, 2004; Herrald & Tomaka, 2002). Imagery, for example, may introduce significant variability in the emotional situations experienced, which may in turn hinder the physiological discrimination between distinct emotions (Rainville et al., 2006). Autonomic changes accompanying emotion are also thought to provide essential support for action, i.e., mobilising energy resources necessary for specific behaviours (Davidson, 1994; Levenson, 2003b). Different behaviours, however, may be necessary or adaptive under diverse circumstances, e.g., fear may be associated with either vigilance or flight, depending on the situation (Lang, Bradley, & Cuthbert, 1990). By this view, physiological arousal responses may also vary depending on the nature of the action tendencies recruited by the emotion in a specific context (see, e.g., Dickerson, Gruenewald et al., 2004). Also consistent with this view is positive emotions' general lack of association with distinct autonomic patterning, which may be attributed to the fact that, unlike negative emotions, they are not strongly associated with specific action tendencies or high-activity motor programs (Fredrickson & Levenson, 1998; Frijda, 1986).

Stemmler and colleagues (2001) argued against absolute emotion specificity, and incorporated the effects of various situational and motivational variables on somatovisceral activation in their component model of somatovisceral response organization. According to this model, multiple influences are superimposed during any emotion elicitation to create the net somatovisceral response pattern; the same emotion may thus lead to different physiological profiles depending on the *context* within which the emotion is induced. They argue that there are three sources that may influence somatovisceral activation during discrete emotional episodes. The first source constitutes the non-emotional context of the induction, including body position and temperature, ongoing motor activity, and cognitive demands of the situation. The second source reflects the emotion signature proper (i.e., noncontextual somatovisceral activation specifically elicited by the emotion in question). The physiological response of a particular emotion is hypothesized to be conserved across different contexts, and serves to protect and prepare the organism for prototypical behavioural responses. The last source of variance is necessitated by the contextual demands of the situation, and includes motivational and behavioural direction in the pursuit of an emotional goal. This component thus allows for flexible modification of bodily resources in accordance with momentary demands.

To determine whether context effects on somatovisceral responses may be separated systematically, Stemmler and colleagues (2007) devised an experiment where they independently manipulated the effects of both emotion and motivational direction. Their paradigm involved a between-subjects script-driven imagery task where participants had to visualize themselves as the protagonists in anger and fear scenarios that differed in motivational direction (i.e., either approach or withdrawal). Results indicated that emotion and motivation contributed independently to somatovisceral effects, and could be separated though statistical analysis.

Taken together, emotions may be associated with specific somatovisceral activation, although the detection of a specific response pattern may be confounded by context effects. Investigating the physiological profile of any emotion, however, remains an important step in unravelling its psychological processes and relationship to physical and mental health. Put differently, response patterns of autonomic activation within sympathetic, parasympathetic, and humoral axes may serve as objective measures to characterize and distinguish internal phenomena that are not readily quantifiable through subjective report.

The autonomic nervous system (ANS). To conclude this section I will describe autonomic arousal responses in the body, focusing on one influential psychophysiological theory that relates autonomic function to social behaviour: the polyvagal theory (Porges, 1995).

The autonomic nervous system constitutes the principal means to regulate the body's internal environment (Brownley, Hurwitz, & Schneiderman, 2000). It is subdivided into a sympathetic (SNS) and parasympathetic (PNS) nervous system, which dually innervate most internal organs. The SNS and PNS, however, usually have antagonistic effects on the target organ: Sympathetic activation prepares the body for action (i.e., "fight or flight") by increasing both cardiac chronotropy (contractile rate) and cardiac inotropy (contractile strength) to facilitate motor action, whereas parasympathetic activation is associated with energy conservation and recuperative actions, such as heart rate slowing and decreased blood pressure. SNS influences on the heart are mediated by the release of norepinephrine from sympathetic nerve terminals. In contrast, PNS deceleration of the heart is mediated by the myelinated vagus nerve, which releases the neurotransmitter acetylcholine.

In psychological research, heart rate reactivity is often conceptualized as a unitary construct that depends only on individual differences in sympathetic reactivity (Cacioppo, Uchino, & Berntson, 1994). Cardiac chronotropy, however, may be influenced by multiple modes of autonomic control, including sympathetic or beta-adrenergic stimulation, vagal withdrawal, or some combination of the two (Berntson, Cacioppo, & Quigley, 1991). Cacioppo et al. (2000) therefore iterated that emotions may be more readily differentiated if contributions from both the SNS and PNS were considered, rather than focusing on visceral responses per se.

The polyvagal theory is based within an evolutionary framework that relates different modes of autonomic responding with adaptive social behaviour (Porges, 1995, 1997, 2001, 2007). Specifically, Porges articulated three autonomic subsystems that are phylogenetically ordered and that respond to challenges in a hierarchical fashion depending on the specific context. First is the social communication system that fosters calm and prosocial behaviours when the environment is perceived as safe. This system is dependent upon the activity of the phylogenetically new and fast-acting myelinated vagus nerve, which originates in the brainstem nucleus ambiguus and terminates on the heart's cardiac pacemaker (the sinoatrial node). Increased efferent activity of the myelinated vagus slows the heart and has an inhibitory effect on cardiac sympathetic influences. The second system is associated with active mobilization to cope with threats, and comprises the sympathetic nervous system. In challenging situations, the vagal brake is rapidly withdrawn, leading to SNS dominance and increased metabolic output. Finally, the immobilization system is phylogenetically the most primitive mode of response, it leads to freezing or feigning death, and it depends on activity of the slow-responding unmyelinated vagus nerve. According to the phylogenetic hierarchy, mammals respond with the newest vagus system first, which inhibits the older systems. When defensive strategies are called upon, however, the older systems are recruited sequentially.

The polyvagal theory has inspired much research in diverse areas of psychological enquiry, and has allowed researchers to formulate testable hypotheses of pathological disease progression in both physical and mental health (e.g., Beauchaine, 2001; Beauchaine, Gatzke-Kopp, & Mead, 2007; Rottenberg, 2007). The theory states that the myelinated vagus is a dynamic contributor to a range of social and emotional processes by way of its ability to rapidly regulate visceral homeostasis (Porges, 2003; 2007). Another central premise of polyvagal theory is that the brainstem nuclei that regulate the myelinated vagus also have afferent fibers to the cranial nerves that mediate facial expression. This bi-directional coupling allows a mechanism by which cardiac states can be coordinated with spontaneous social behaviours.

Estimates of cardiac vagal activity are usually obtained through unobtrusive measures of heart rate variability (HRV) (Berntson et al., 1997). HRV may be described as rhythmic oscillations in the time between consecutive heartbeats, with greater variability considered to reflect greater behavioural flexibility in response to environmental demands (Porges & Byrne, 1992). While the two autonomic branches dually modulate the heart's interbeat intervals, they rely on different signalling mechanisms with different temporal characteristics, which in turn form the basis for HRV frequency analysis (Montano et al., 2009). SNS influences mediated by norepinephrine unfold and decay rather slowly (4 s to 20 s), whereas PNS influences mediated by acetylcholine have a very short response latency (0.5 s to 1 s). Only PNS regulation of the heart via the vagus, therefore, is capable of modulating beat-to-beat cardiac activity. The rhythmic variation of cardiac rate in response to respiration, known as respiratory sinus arrhythmia (RSA), is viewed as a relatively pure estimate of vagal activity because only PNS modulation of the heart is fast enough to covary with respiration (Berntson et al., 1997). RSA results from the rhythmic gating of vagal efference during breathing: during inhalation vagal outflow is reduced and HR accelerates, and during exhalation vagal outflow is increased and HR decelerates. It should be noted, however, that the use of RSA as a measure of cardiac vagal control may be confounded by respiratory frequency and depth, which should ideally be taken into account in experimental designs (Allen, Chambers, & Towers, 2007; Grossman & Taylor, 2007).

To calculate HRV, a continuous electrocardiogram (ECG) of heart rate is required that permits reliable identification of R-R intervals (see Allen et al., 2007 for details). HRV may then be assessed via either statistical analyses (time domain) or frequency domain measures (Task Force, 1996). To obtain frequency measures, power spectral analysis is employed to partition the amount of HRV occurring at different frequencies into a power spectrum. Prominent bands on this power spectrum represent the major oscillatory components of HRV, with the two most reliable periodicities occurring at a high frequency (HF) and a low frequency (LF). The HF (.15-40 Hz) component occurs at the frequency associated with spontaneous adult breathing, and primarily reflects PNS control over the heart due to respiratory sinus arrhythmia (Hayano et al., 1991).

Controversy exists, however, with regard to the physiological origin of the LF component (.04-.15 Hz). While some contend that LF heart rate variability reflects primarily cardiac sympathetic innervation (Malliani, Pagani, Lombardi, & Cerutti, 1991; Pagani et al., 1991), most researchers would agree that the LF component also contains parasympathetic influence, and thus reflects a combination of SNS and PNS activity (Eckberg, 1997; Martinmäki, Rusko, Kooistra, Kettunen, & Saalasti, 2006). The LF component is also assumed by some to be related to the endogenous rhythm of baroreceptor activity, which functions to regulate blood pressure (Moak et al., 2009; Sleight et al., 1995). When blood pressure is sympathetically elevated in the periphery, baroreceptors are activated that stimulate a central reflex mechanism in the brainstem to inhibit SNS activation and stimulate vagal efference, thereby lowering blood pressure. LF spectral power may therefore also reflect baroreflex-mediated vagal modulation.

To conclude, HRV may be viewed as a noninvasive experimental measure of physiological flexibility. Individuals with higher resting heart rate variability are thought to be more emotionally responsive to the environment and less likely to be associated with psychopathology and cardiovascular disease (Beauchaine, 2001; Thayer & Lane, 2007). HRV may also serve as a proxy for more central regulatory processes (Porges, 1996; Thayer & Lane, 2000). HRV, for example, has been used as an estimate of sustained attention (Weber, Van Der Molen, & Molenaar, 1994), and an index of emotional stress and stress vulnerability (Porges, 1995). More recently, HRV has also been associated with individual differences in emotion regulatory ability (Appelhans & Luecken, 2006; Butler, Wilhelm, & Gross, 2006). A growing body of literature thus points to the prominent involvement of HRV in emotional responding, yet more research is necessary to elucidate its precise function.

Neurobiology of Moral Emotions

Moral behaviour is fundamentally important to human beings: It forms the bedrock of an ethical society and allows us to flourish in our rich social environments. Yet relatively little is known about the neural underpinnings of moral cognition (Casebeer, 2003). Over the past decade, a growing number of researchers have started to use functional neuroimaging and clinico-anatomical methods to uncover the neural substrates of moral behaviour. Despite these efforts, we still know relatively little about how moral judgments, as well as moral emotions, are instantiated in the brain, and what the particular contributions of various neural regions may be. This state of affairs may be attributed to numerous factors, including the highly distributed neural architecture of moral cognition (Casebeer, 2003), the lack of ecological validity of most functional imaging paradigms (Levenson, 2003a; Moll, Zahn et al., 2005), and limitations of neuroimaging methods themselves.

Neuroimaging techniques such as functional magnetic resonance imaging (fMRI) and positron emission tomography (PET) detect changes in local cerebral blood flow during mental activities (Raichle, 1994). Whereas PET measures local variations in regional cerebral blood flow (rCBF) through use of a radioactive tracer, fMRI measures the ratio of oxygenated to deoxygenated hemoglobin in active parts of the brain, known as the blood-oxygenation-level-dependent or BOLD signal (Ogawa, Lee, Kay, & Tank, 1990; Raichle, 1998). Both techniques are thus indirect measures of neuronal activity, centered on the principle that areas with increased activity will have a corresponding increase in local blood flow. Although there is reason to believe that a high degree of coupling exists between neural activity and local blood flow changes (neurovascular coupling), the exact nature of this relationship is not known (Arthurs & Boniface, 2002; Logothetis, 2003; Logothetis & Pfeuffer, 2004). Moreover, the temporal and spatial resolution of fMRI, and especially PET, is limited (Aguirre, Zarahn, & D'Esposito, 1998).

Despite these limitations in inferring mind-brain relationships through functional imaging techniques, findings from imaging studies on moral cognition consistently point to the involvement of the same neural networks (Moll, Zahn et al., 2005). Both neuroimaging and clinical evidence have implicated the involvement of partially dissociable prefrontal, temporal, and limbic networks in moral phenomena (Allison, Puce, & McCarthy, 2000; Grafman et al., 1996; Kiehl et al., 2001; Miller, Chang, Mena, Boone, & Lesser, 1993; Muller et al., 2003; Sturm et al., 2006; Tranel, 1994). Neocortical, as well as phylogenetically older mesolimbic and orbitofrontal (OFC) regions, are therefore involved. The moral emotions are complex subjective experiences and are also thought to arise from distributed activation in these areas (Zahn, Moll et al., 2009). It is still uncertain, however, whether the neural circuitry recruited during moral emotion elicitation is similar across different emotions, or whether it differs depending on the specific moral emotion in question (Kédia et al., 2008). At this stage, available evidence suggests that both qualitative and quantitative differences exist between the neural correlates of distinct moral emotions (Takahashi et al., 2004; Zahn, Moll et al., 2009).

Neurobiology of basic versus moral emotions. fMRI studies that investigate brain activation associated with basic emotions, e.g., the positive emotion of happiness, and especially negative emotional states, like sadness, fear, anger, anxiety, and disgust, typically report increased activation in limbic and paralimbic structures (Damasio et al., 2000; George et al., 1996; George et al., 1995; Kimbrell et al., 1999; Lane, Reiman, Ahern, Schwartz, & Davidson, 1997; Pardo, Pardo, & Raichle, 1993; Phillips et al., 1997; Reiman et al., 1997). These activations are readily explained when one views the basic, more survival-laden, emotions as powerful motivational states that are linked to the brain's regulatory core (Casebeer, 2003).

Brainstem and limbic areas not only monitor bodily homeostasis, but also appear to underlie the instantiation of central motivational states and the processing of affectively significant information (Mesulam, 1985; Stellar, 1994). Damasio et al. (2000) showed that brainstem tegmentum, hypothalamic, and limbic activations were associated with several basic emotions, thus underscoring the essential relation between structures involved in the representation/regulation of the body's internal milieu and viscera, and subjective feeling states. Moreover, Damasio and colleagues suggested that the distinct neural patterns associated with each emotion may form the basis for the mental states known as feelings (Damasio, 1999; Damasio et al., 2000). It is important to note that basic emotions, like fear and sadness, can be distinguished from central motivational states, in that basic emotions consist of the temporal binding of contextual information (e.g., perceiving a feared object) and the central motivational state itself (e.g., undirected anxiety) (Moll, Zahn et al., 2005).

Basic emotions have been associated with substantial variability in terms of their neural activation patterns. Although this variation may reflect the large range of elicitation methods and stimuli employed in basic emotion experimental paradigms (Moll, Zahn et al., 2005), it also supports the notion that different emotional states engage brain regions in unique patterns. Different limbic structures are typically associated with different basic emotions, and are thought to represent the specific emotional quality (Paradiso et al., 1997; Zahn, Moll et al., 2009). For example, sadness has consistently been associated with anterior paralimbic structures, e.g., ventromedial PFC/subgenual cingulate (George et al., 1995; Lane, Reiman et al., 1997; Mayberg et al., 1999); anger has been associated with activation in the orbitofrontal cortex and anterior cingulate areas (Damasio et al., 2000; Dougherty et al., 1999); disgust typically activates the anterior insula (Phillips et al., 1997; Wright, He, Shapira, Goodman, & Liu, 2004); and fear has consistently been associated with the amygdala (LeDoux, 2003; Morris et al., 1996).

Human moral cognition, on the other hand, is considered intimately entwined with the evolution of the prefrontal cortex (Wood & Grafman, 2003). The PFC, and other evolutionally 'newer' forebrain areas, remained connected to and are innervated by the brain's regulatory core; limbic and paralimbic structures therefore also serve as a foundation on which the moral emotions are built (Casebeer, 2003). The anterior cingulate cortex, for example, is a large paralimbic structure that plays an integral role in the interface between cognition and emotion (Allman, Hakeem, Erwin, Nimchinsky, & Hof, 2001). Of primary importance, however, is the PFC, which is regularly activated by moral emotion imaging studies (Moll, de Oliveira-Souza, Eslinger et al., 2002), and has been proposed to form part of a complex neural circuit involved in the instantiation and regulation of emotion and motivation (Davidson, Pizzagalli, Nitschke, & Kalin, 2003). Ventromedial and orbitofrontal PFC, in particular, are regarded as important neural substrates in the automatic processing of moral emotions, because dysfunction in these areas can result in inappropriate or inefficient moral emotions (Beer et al., 2003; Koenigs et al., 2007), reduced empathy (Eslinger, Parkinson, & Shamay, 2002), and antisocial behaviours (Saver & Damasio, 1991). In comparison, the rostral frontopolar cortex appears to be more involved in explicit moral reasoning and prospective assessments (Greene et al., 2001; Moll, Eslinger, & Oliveira-Souza, 2001; Okuda et al., 2003), as well as prosocial behaviour (Masten et al., 2011).

Compared to the 'less complex' basic emotions, a remarkable overlap in brain activation is observed in moral cognition imaging studies. In a review of moral cognition, Moll and colleagues (2005) highlighted a network of structures that were consistently found to be active across a wide range of stimulus modalities and task requirements. Notably, these structures were activated in studies employing moral judgment/reasoning tasks (e.g., Greene et al., 2001; Heekeren, Wartenburger, Schmidt, Schwintowski, & Villringer, 2003; Moll, de Oliveira-Souza, Bramati, & Grafman, 2002), as well as studies investigating specific moral emotions (e.g., Berthoz et al., 2002; Moll, de Oliveira-Souza, Eslinger et al., 2002; Takahashi et al., 2004). Brain structures activated included the anterior PFC (BA 9/10), medial and lateral orbitofrontal cortex (OFC; BA 10/11/25), medial and ventromedial PFC (BA 12/32), the anterior temporal lobes (BA 20/21/38), the superior temporal sulci (STS; BA21/39), insula, anterior cingulate cortex (ACC; BA 24/32), precuneus (BA 7/31), and limbic regions. Moll et al. (2005) argued that moral phenomena (including moral emotions) rely on the binding of three main components: Context-dependent representations of events (i.e., structured event knowledge) provided by prefrontal regions, social perceptual or context-independent featural representations (i.e., semantic properties extracted from different social situations), which are stored in the postero-superior and anterior temporal regions respectively, and central motivational/emotional states (e.g., aggression, sexual desire, anxiety, sadness, hunger, attachment), which are represented in limbic and paralimbic structures. When these components interact and are activated together by way of temporal synchronization, they give rise to a 'gestalt' experience (Singer, 2001). For example, during compassion, cortical representations may allow one to notice the emotional need of someone in distress, while central motivational states will give rise to feelings of sadness, attachment and anxiety, which will move one to help the person in pain (Moll et al., 2005). These component representations are consistently detected in different moral emotion studies.

Of particular interest is the fact that the structures activated by moral cognition paradigms include areas most commonly associated with theory of mind (ToM) ability, namely the medial prefrontal cortex (mPFC), the superior temporal sulci (STS) and temporoparietal junction (TPJ), and the temporal poles bilaterally (Frith & Frith, 1999; Gallagher & Frith, 2003). As explained previously, the ability to represent the mental states of others, i.e., ToM, appears to be integral to the experience of self-conscious emotions (Eisenberg, 2000; Heerey et al., 2003; Leary, 2004, 2007). Robust ToM is in fact necessary for a whole range of morally important cognitive abilities (Casebeer, 2003). For example, a negative self-conscious emotion may require the recognition that a social norm has been broken, as well as the negative evaluation of self, both of which are important ToM abilities (Takahashi et al., 2004).

In support of the notion that self-conscious emotions are intimately tied to inferences about others' evaluations of self, several imaging studies on these emotions have found activations in regions implicated in ToM (e.g., Berthoz et al., 2002; Takahashi et al., 2004; 2008). The importance of ToM ability in the production of complex social emotions was even more evident in Kedia et al.'s (2008) study, where they examined the distinct moral emotions that result when 'an agent harms a victim'. By varying the involvement of self and other as either the agent or victim of a moral transgression, they created scenarios that evoked four different moral emotions: self-anger ("I harm myself), guilt ("I harm someone else"), other-anger ("someone else harms me"), and compassion ("someone else harms himself"). Kedia et al.'s results indicated that only conditions that also involved other people (all conditions except that of self-anger) activated structures that are known to be involved in ToM, including the dorsal mPFC, precuneus, and bilateral TPJ. These areas are thought to play unique roles in the brain's ability to perceive and reason about other people (Saxe, 2006; Saxe & Kanwisher, 2003). Other regions commonly implicated in the ToM network include the posterior cingulate cortex and limbic/paralimbic structures, such as the amygdala and orbitofrontal cortex (Fletcher, Happe et al., 1995; Gallagher & Frith, 2003; Greene & Haidt, 2002).

Because the amygdala is intimately associated with emotion in general, a closer look at its role in moral emotion is appropriate. As mentioned above, the amygdala is most consistently associated with the capacity to generate/recognize fear-related emotions and detect threat (Amaral, 2003; Breiter et al., 1996; LeDoux, 2000). It has also, however, been associated with the processing of positive/interesting stimuli, e.g., happy faces and unusual pictures, and appears to be involved in processing the salience of emotional stimuli in general (Davis & Whalen, 2001; Hamann & Mao, 2002; Liberzon et al., 2003; Somerville, Kim, Johnstone, Alexander, & Whalen, 2004). While the amygdala is thus commonly implicated in the processing of (salient) basic emotions, Adolphs (2003) has suggested that the amygdala's function in humans evolved to play a specific role in detecting subtle social signals in everyday life. By way of support for this notion, Adolphs and colleagues found that patients with either unilateral or bilateral amygdala lesions showed worse impairments for the perception of social emotions, than that of basic emotions (Adolphs, Baron-Cohen, & Tranel, 2002). Consistent with this interpretation, amygdala activation has been observed in imaging studies of various social emotions, e.g., other-anger (Grèzes, Berthoz, & Passingham, 2006), empathy (Völlm et al., 2006), and guilt (Berthoz, Grèzes, Armony, Passingham, & Dolan, 2006). It is therefore unlikely that, in humans, the amygdala only plays a role in the more survival-laden basic emotions, although clinical evidence for severe moral behavioural dysregulation in patients with focal amygdala lesions is still missing (Moll, Zahn et al., 2005).

Review of moral emotion neuroimaging studies. While several neuroimaging studies in the past decade have examined the neural correlates associate with various complex social emotions, e.g., jealousy, compassion, romantic and maternal love, admiration, indignation, and social pain (Aron et al., 2005; Bartels & Zeki, 2004; Eisenberger & Lieberman, 2004; Immordino-Yang, McColl, Damasio, & Damasio, 2009; Moll, de Oliveira-Souza et al., 2005; Takahashi et al., 2006), only a handful of these focused specifically on guilt. The first was a PET study conducted by Shin et al. (2000), who employed a script-driven imagery paradigm to study regional changes in cerebral blood flow associated with reliving a previous guilty episode. Reliving guilt, compared to a neutral condition, was associated with enhanced neural activity in several anterior paralimbic areas, including the anterior cingulate cortex, left anterior insula and inferior frontal gyrus, and the temporal poles. Almost all subsequent paradigms made use of either short sentences or vignettes depicting moral violations to elicit guilt. I provide a brief summary of these studies below.

In one of the first fMRI studies that used emotive sentences to study moral phenomena, Berthoz et al. (2002) investigated the neural responses to stories depicting both intentional and unintentional (embarrassing) violations of social norms, with either a personal or impersonal reference. They thus aimed to study the effects of intentionality and agency in social violations. In response to both types of social violations, Berthoz et al. reported increased activity in areas associated with ToM, as well as in areas associated with processing aversive emotional expressions (e.g., lateral OFC, BA 47). They argued that the ToM activations may have been associated with attempts to infer the intentions of the offender, while the OFC activation may have corresponded to the processing of actions that evoke anger in other people. In a subsequent publication, they reanalyzed their fMRI data and found that intentional social violations performed by the self, which may be associated with guilt, were additionally associated with bilateral amygdala activation (Berthoz et al., 2006).

Similarly, Takahashi et al. (2004) assessed the evaluative processes of guilt and embarrassment through an emotional judgment task. Their participants thus rated the described statements according to how guilty/embarrassing they appeared, e.g., Guilt: "I betrayed my friend"; Embarrassment: "I noticed that the zipper of my pants was open." Takahashi et al.'s (2004) fMRI results also implicated ToM areas for both emotion conditions, while the embarrassment condition additionally activated areas such as the right temporal cortex and the bilateral hippocampi.

Using a slightly different approach, Finger et al. (2006) studied the effects of the presence or absence of an audience on neural responses to moral and social transgressions. They employed short vignettes in the second person (i.e., "You did this"), and argued that moral transgressions with or without an audience should be associated with the experience of guilt, whereas only witnessed social transgressions should be associated with embarrassment (Keltner & Buswell, 1997). Besides activation in ToM areas (mPFC, bilateral temporal poles, and left STS), Finger et al. reported that the dorsomedial (BA 8) and ventrolateral (BA 47) PFC showed increased BOLD responses in all conditions that required a change from the current behaviour, i.e., a change away from the 'bad' behaviour. These conditions included moral transgressions regardless of whether an audience was present, but only social transgressions in the presence of an audience.

The importance of agency as a major dimension in moral emotion was further explored by Moll et al. (2007), who independently addressed agency and emotional processing in several moral emotions. While the sense of agency engaged ventral frontal and temporal areas, prosocial emotions (including guilt, embarrassment and compassion) were associated with activation of the anterior mPFC and STS/temporoparietal junction (TPJ). Empathic emotions (guilt and compassion) additionally activated the mesolimbic reward pathway (e.g., ventral tegmentum/thalamus), which plays an important role in social attachment (Insel & Young, 2001). The guilt condition was also associated with activation in the left anterior insula. In the Kedia et al. (2008) study described above, the guilt and other-anger conditions also activated emotional structures, including the bilateral amygdala, ACC and basal ganglia, in addition to ToM areas.

In a continued effort to disentangle important context variables of moral emotions, Moll and colleagues designed a study in which agency (self versus other), as well as value-related actions, were manipulated (Zahn, Moll et al., 2009). Participants thus imagined themselves (self-agency), or someone else (other-agency), performing actions that were either in accordance with a social value (resulting in pride or gratitude, respectively), or that were counter to a social value (resulting in guilt or anger/indignation, respectively). Zahn et al.'s results supported the notion that abstract social knowledge is represented within anterior temporal regions, while distinct moral emotions engaged differential activations within fronto-mesolimbic regions. Specifically, the individual difference effect for guilt (i.e., higher individual frequency of guilt) was associated with increased activity in the subgenual cingulate cortex (SCC, BA 32) and anterior ventromedial PFC. In a subsequent study and reanalysis of fMRI data, however, Zahn and colleagues reported that the subgenual cingulate has a direct association with empathic concern, rather than guilt *per se* (Zahn, de Oliveira-Souza, Bramati, Garrido, & Moll, 2009). The pride condition was associated with activation within the mesolimbic reward pathway, e.g., the ventral tegmental area, and higher individual frequency of pride was specifically associated with increased activity in the tectum. Another fMRI study that specifically examined pride, however, only reported increased activity in temporal areas (Takahashi et al., 2008).

Finally, a recent study employed a novel paradigm based on the alternating presentation of emotive facial expressions and contextual sentences to examine neural activations associated with guilt (Basile et al., 2011). Specifically, Basile and colleagues aimed to differentiate between guilt driven by internal values, e.g., breaking a personal religious rule (deontological guilt), and guilt that arises more in the context of interpersonal situations, e.g., feeling guilty about someone else's bad luck (altruistic guilt). Results for the main effect of guilt indicated significant activation in the anterior and posterior cingulate, and in the left medial prefrontal gyrus. Deontological guilt, however, was associated with activation in a more dorsal ACC area (supragenual cingulate), while altruistic guilt was associated with activation in a more ventral area within the left mPFC. In addition, the left anterior insula responded selectively to deontological guilt, suggesting that this form of guilt entails aversive emotional processing (Phan, Wager, Taylor, & Liberzon, 2002).

Taken together, the information presented above suggests that the instantiation of moral emotions are associated with structures involved in cognitive/emotion integration and higher-order context-specific representations (i.e., PFC and ACC), areas involved in monitoring one's own and another's mental state (ToM) (i.e., mPFC, temporal areas, posterior cingulate, and precuneus), and areas associated with emotional processing, i.e., limbic regions (Greene & Haidt, 2002; Moll et al., 2007; Moll, Zahn et al., 2005). Guilt, in particular, appears to be associated with activation in the ACC extending to the medial prefrontal cortex, as well as the anterior insula (particularly on the left) (Basile et al., 2011; Kédia et al., 2008; Moll et al., 2007; Shin et al., 2000; Zahn, Moll et al., 2009). While prefrontal areas are necessary for higher-order cognitive processes, the anterior cingulate and anterior insular cortices are considered essential in forming an emotion (Craig, 2002). The insula is specialised for mapping visceral responses and generating subjective feeling states (limbic sensory cortex), in turn, the ACC is associated with behavioural motivation because of its association with autonomic and emotional control (limbic motor cortex) (Craig, 2002; Critchley, 2005). Together, they thus represent emotional feeling and drive, respectively. The association of these areas with guilt also concurs with findings from patients with frontotemporal lobar degeneration. Such patients are characterized by diminished self-conscious emotional responding, which is thought to be the result of selective neural loss in frontal and insular brain regions (Sturm et al., 2006, 2008).

Another important conclusion that may be drawn from the literature reviewed above, is that the emotion elicitation method or study design will probably impact significantly on the neural activations observed. For example, real-life experiences of moral emotions are expected to be far more emotionally arousing than emotions elicited through descriptive sentences or vignettes (Finger et al., 2006). One would therefore also expect to see increased neural activation in areas associated with increased emotional arousal in response to real-life emotional events, e.g., the amygdala, insula, anterior cingulate, and medial OFC (Critchley, 2005; Critchley, Mathias, & Dolan, 2002; Phelps & LeDoux, 2005). In a similar vein, George et al. (1996) argued that the regions activated during actual sadness may be different from remembering emotionally-laden memories or seeing sad human faces. Even more intriguing is preliminary evidence suggesting that *externally* generated emotions (e.g., those generated by visual stimuli) involve the amygdala to a greater extent than *internally* generated emotions (e.g., memories of personal emotional events) (Damasio et al., 2000; Reiman et al., 1997). Data derived from the meta-analysis conducted by Phan et al. (2002) indicated that emotion elicitation through use of visual aids (e.g., pictures or faces) readily activated the amygdala, whereas emotion elicitation through recall/imagery tasks tended to activate the anterior cingulate and insula. The amygdala thus appears to be specialized for processing visually relevant emotional cues in order to alert us to threats in our environment (Zald, 2003).

A final consideration drawn from the literature reviewed above involves the importance of the self as the moral agent in moral emotion elicitation studies. Various fMRI studies of moral dilemmas have reported increased activity in affective structures when one is personally involved in a moral situation (Greene et al., 2004; Greene et al., 2001; Kédia et al., 2008). For example, when the self is directly responsible for the harm done to someone else (e.g., you have to push a person in front of a trolley in order to save five others), increased activity in the amygdala and basal ganglia is observed, compared to when your actions have more indirect effects (e.g., by hitting a switch you can divert the course of the trolley to kill one person, but in the process save the lives of five others) (Greene et al., 2004). Further support for the importance of the direct involvement of the self in moral emotion elicitation comes from recent fMRI studies, where enhanced activity in the amygdala was observed when participants were involved as either the agent or victim of a social transgression, rather than being merely a witness to the scenario (Berthoz et al., 2006; Grèzes et al., 2006).

Notwithstanding these general findings, individual variability among participants in terms of personality and dispositional affect has been shown to impact significantly on brain activation patterns (Canli et al., 2001; Hamann & Canli, 2004). Assuming homogenous responses across participants may thus not accurately reflect the neural bases of emotional processing, especially given that individual differences in the domain of emotion appear to be the rule rather than the exception (Eugene et al., 2003).

Behavioural Motivation: The BIS/BAS Model

In recent years, a central question in neurocognitive research has concerned how individual differences in neurobiological processes relate to personality, motivation and behaviour, and by extension, psychological dysfunction (Curtis & Cicchetti, 2007; Davidson et al., 2000; Harmon-Jones, Gable, & Peterson, 2009; Jackson et al., 2003; Larsen & Prizmic, 2004). Considerable variability exists among healthy adults in the intensity, expression, and regulation of an emotional response, and it is well documented that the intensity of this affect may be predicted from previously obtained subjective measures (Carver & White, 1994; Ochsner & Gross, 2005). The approach I adopted in the current research, therefore, was aimed at evaluating guilt (and pride) in terms of individual differences in affective style, particularly in relation to behavioural motivation tendencies. This is of particular significance, given the conflicting accounts in the literature on guilt's behavioural motivation: Is it an inhibitor of transgressive behaviour, or a promoter of prosocial behaviour (Baumeister et al., 1995; Monteith, 1993; Monteith, Ashburn-Nardo, Voils, & Czopp, 2002; Schmader & Lickel, 2006; Sheikh & Janoff-Bulman, 2010; Tangney, 1991; Tangney & Dearing, 2002)?

One dominant theory of behavioural motivation suggests that there are two fundamental brain systems that govern adaptive behaviour: the behaviour inhibition system (BIS), and the behaviour activation system (BAS) (Fowles, 1980, 1988; Gray, 1975, 1982, 1987b). Following this theory, stable, trait-like individual differences in BIS and BAS sensitivities to emotionally salient stimuli in the environment shape emotional learning and define individual temperament. Because Gray's work was originally derived from work with nonhuman animals (Gray, 1972), the BIS/BAS dimensions of personality have a strong link with neural physiology, unlike most other theories of temperament (Morgan, 2006; Smillie, Pickering, & Jackson, 2006).

The BIS is an attentional system that is sensitive to cues of punishment, extreme novelty, and frustrative nonreward (e.g., situations where continued goal-directed behaviour is not rewarded). In a more recently updated model, Gray and McNaughton (2000) proposed that the BIS is sensitive to any stimuli that generate competing responses tendencies and thus require conflict-monitoring (see also McNaughton & Corr, 2004). Activation of the BIS in response to such conflicts is associated with increased attention and arousal, and functions to momentarily inhibit or halt ongoing behaviour, but also elicits behaviours aimed at resolving the conflict, e.g., environmental scanning and risk assessment (McNaughton & Corr, 2004).

Gray held that BIS functioning supports the experience of negative feelings such as fear, anxiety, frustration, and sadness, with strong BIS corresponding to anxiety-related disorders (Fowles, 1988; Gray, 1981). Gray thus proposed the BIS as the causal basis of anxiety. Neurobiologically, the BIS is associated with the amygdala and septohippocampal system, including both its monoaminergic brainstem afferents and related structures, as well as its neocortical projections to the frontal lobes (Gray, 1982; Gray & McNaughton, 2000). In particular, because the ACC is strongly associated with conflict-monitoring, as supported by numerous studies in cognitive neuroscience (Botvinick, Cohen, & Carter, 2004; van Veen, Cohen, Botvinick, Stenger, & Carter, 2001), BIS activation is thought to act via the ACC to detect conflict and interrupt action (Amodio, Master, Yee, & Taylor, 2008).

The second system, i.e., BAS, is a motivational system that responds strongly to signals of reward, nonpunishment, and escape from punishment. As such, the BAS subserves motivational functions that govern both approach behaviours (i.e., behaviours that maximize reward by moving toward goals) and active avoidance (i.e., behaviours that minimize punishment by moving away from threats). BAS functioning is associated with the experience of positive affect, such as hope, elation, and happiness, but also aggression and anger (Gray, 1981; Harmon-Jones, 2003; Wingrove & Bond, 1998). Extreme levels of BAS have been associated with impulsivity (Gray, Owen, Davis, & Tsaltas, 1983). The neural basis of the BAS is rooted in the dopaminergic neurotransmitter system and includes structures such as the ventral tegmentum, nucleus accumbens, and ventral striatum (Gray, 1987a).

Taken together, the BIS is thus associated with stopping, while the BAS is associated with going. According to Gray, individuals with high BIS sensitivity should be naturally inclined to fixate on possible threats or punishment in their environment, and should be prone to experience negative affect. Conversely, individuals with high BAS sensitivity should be more responsive to cues of reward and more prone to experience positive affect.

Since Gray's original postulations, considerable effort has been made to assess BIS/BAS reactivity in individuals across several response systems, e.g., self-report, physiological, and neuronal. Despite these efforts, however, there is currently no consensus on the appropriate level of analysis for assessing BIS and BAS sensitivity in humans (Corr, 2001). For example, a large body of literature employing electroencephalography (EEG) has provided general support for the association between greater left-sided frontal asymmetry and BAS activation, whereas the BIS does not appear to be directly related to frontal EEG asymmetry (Coan & Allen, 2003a).

One of the most used and best validated self-report scales developed to assess BIS and BAS sensitivity are the Carver and White BIS/BAS scales (Carver & White, 1994). The BIS and BAS, as measured by these scales, are associated with sensitivity to punishment and reward, respectively. Notably, the BAS scale specifically focuses on motivation toward goals and does not include avoidance behaviours, e.g., escape from punishment or threat. It therefore only measures approach aspects of the BAS. Some ambiguity has also been associated with the specific interpretation of the BIS, i.e., whether it is associated with the inhibition of action, as originally proposed by Gray, or whether it reflects behavioural avoidance in response to a threat (Amodio, Master et al., 2008). In a recent study employing EEG measures, Amodio and colleagues (2008) determined that BIS (as measured by the Carver and White BIS/BAS scales) was associated with conflict-related activity located in the dorsal ACC (Nieuwenhuis, Yeung, van den Wildenberg, & Ridderinkhof, 2003). In their study, self-reported BIS scores were uniquely related to successful response inhibition, as assessed by the N2 event-related potential on No-Go trials in a Go/No-Go task, and unrelated to approach/avoidance tendencies, as assessed by baseline frontal cortical asymmetry. They interpreted these findings as providing support for BIS's association with behaviour inhibition rather than avoidance.

A substantial body of research has also examined the BIS and BAS in relation to peripheral measures of physiological activity during experimental manipulations of reward or punishment, in an effort to associate specific measures of cardiac reactivity and electrodermal activity with BIS and BAS reactivity (Arnett & Newman, 2000; Fowles, 1980, 1988; Gomez & McLaren, 1997; Knyazev, Slobodskaya, & Wilson, 2002). In this psychophysiological tradition, physiological correlates of BIS and BAS are typically assessed in situations of appetitive responding for reward (i.e., BAS), and during frustrative nonreward or extinction (i.e., BIS). Although this approach has revealed several putative physiological markers of behavioural approach and behavioural inhibition activation, results have not always replicated well across different studies and paradigms (e.g., Heponiemi, Keltikangas-Järvinen, Kettunen, Puttonen, & Ravaja, 2004; Knyazev et al., 2002). Moreover, self-report measures of BIS/BAS sensitivity generally appear to show only a weak relation to physiological measures of reactivity (Brenner, Beauchaine, & Sylvers, 2005).

Despite these inconsistencies, BAS activation has consistently been associated with HR acceleration during tasks of reward, while BIS activation is most consistently associated with electrodermal activity during frustrative nonreward (Fowles, 1980, 1988; Fowles, Fisher, & Tranel, 1982; Tranel, 1983). Conflicting accounts in the literature, however, exist on which peripheral index (i.e., the sympathetic or parasympathetic nervous system) mediates the relationship between BAS and HR reactivity. While some studies have found significant correlations between self-reported BAS and parasympathetic withdrawal, as indexed by RSA (Heponiemi et al., 2004; Knyazev et al., 2002; Ravaja, 2004), others have found SNS-linked cardiac reactivity, as indexed by PEP, to be a superior physiological estimate of approach motivation (Brenner et al., 2005; Crowell et al., 2006). Different findings might be a result of different tasks used, different quantification of behavioural activation/inhibition tendencies, and the presence or absence of incentives, across different studies.

Based on psychophysiological evidence and the Polyvagal Theory, however, Beauchaine (2001) has argued that both the BIS and BAS should be mediated by the sympathetic nervous system. This assumption is derived from numerous observations indicating that emotional lability and psychopathology are associated with reduced cardiac vagal tone and excessive vagal reactivity, resulting in sympathetically-mediated fight/flight response strategies (i.e., approach or avoidance) (Beauchaine et al., 2007). Based on functional and phylogenetic considerations, Beauchaine and colleagues have asserted that PEP reactivity marks behavioural approach/active avoidance (i.e., BAS), while electrodermal responding is associated with behavioural inhibition (i.e., BIS) (Beauchaine, 2001; Beauchaine, Katkin, Strassberg, & Snarr, 2001). Conclusive data confirming these assumptions, however, remain elusive.

SPECIFIC AIMS AND RATIONALE

The overarching goal of this thesis was to study the manifestation of guilt in the body and in the brain by investigating the autonomic and neurophysiological responses to current experiences of guilt. By investigating guilt in different response systems, i.e., by measuring subjective emotional experience, autonomic activation, and haemodynamic brain activity in response to guilt-evoking stimuli, I hoped to perform integrative research that would advance our understanding of how these response systems are coordinated in an emotional response (Larsen & Fredrickson, 1999). Such an interdisciplinary approach may also be well suited to inform our current understanding of how this important moral emotion regulates social behaviour.

In order to realize this goal, however, I required emotion elicitation paradigms that would be effective in the different settings employed, i.e., in the laboratory as well as the MRI scanner. Although ecologically valid emotion elicitation procedures are particularly difficult to design for the moral emotions (Casebeer, 2003), my Literature Review highlighted the importance of such an approach. In a nutshell, ecological validity entails the design of emotion elicitation procedures that evoke "real" emotions that may be representative of emotional events participants encounter in real life, and that can be distinguished from situations where participants merely make emotional judgments (Levenson, 2003a).

My first aim, therefore, was to develop an elicitation paradigm to investigate the real-time physiology of guilt during a controlled experimental situation. A critical aspect of the method employed was that it should afford participants with a personal sense of agency, so that they would perceive themselves as being responsible or playing a role in the harm caused to another (Kubany & Watson, 2003). Based on Harmon-Jones and colleagues'(2007) recommendations, I considered a social psychology manipulation that is high in realism and that involves deception to be the most effective approach. As a contrasting self-conscious moral emotion, I also investigated pride, based on Cacioppo et al.'s (1993) suggestion that emotion-specificity researchers should include positive and negative emotions, as well as neutral conditions for baseline comparison, in their study designs.

For Study 1 I therefore designed and tested a novel social psychology paradigm to examine the physiological profiles of guilt and pride. I aimed to describe the autonomic responses in terms of both parasympathetic and sympathetic nervous system contributions, and also interpreted results in terms of individual differences in sensitivity to Gray's behaviour activation system (BAS) and behaviour inhibition system (BIS) (Gray, 1982, 1987b; Gray & McNaughton, 2000).

My second aim was to investigate the neural correlates associated with current guilt through functional magnetic resonance imaging (fMRI). Once again, I was compelled to develop a novel paradigm of high ecological validity that would be suitable to elicit guilt within the MRI chamber. This time, the paradigm was designed to evoke guilt in low-prejudice individuals as a result of bogus feedback on an implicit measure of prejudice, namely the Implicit Association Test (IAT). The guilt that was elicited by this manipulation may be described as deontological rather than altruistic, because it involved the violation of a personal moral norm, rather than an interpersonal situation in which someone else was harmed (Basile et al., 2011). I again included pride as a contrasting emotion condition for comparison purposes.

My final goal was to integrate results from Study 1 and 2 to provide a holistic account of guilt's effects in the body, as well as its behavioural and motivational functions in everyday life.

STUDY 1: THE PSYCHOPHYSIOLOGY OF GUILT

Investigating the physiology of discrete emotions serves several important functions, including uncovering emotion-specific physiological activity (Levenson, 2003a), and investigating the relation between affect and specific health outcomes (e.g., Lerner et al., 2007; Steptoe & Brydon, 2009). Notably, such investigations also help inform psychological theories of emotion with regard to their motivational and behavioural functions (Amodio, Devine, & Harmon-Jones, 2007; Cacioppo & Gardner, 1999; Izard & Ackerman, 2000). Compared to investigations into the basic emotions (e.g., Rainville et al., 2006), however, the psychophysiological exploration of moral emotions has received little attention to date. The dearth of such literature may be attributed largely to inherent methodological challenges associated with the successful elicitation, as well as the measurement, of moral emotions (Lewis, 2000; Tangney, 1996).

Moral emotions are internal affective states that are linked to the wellbeing of other individuals or society as a whole (Tangney et al., 2007). As early as 1884, William James argued that not only basic emotions, but also the moral emotions, are associated with physiological arousal. He maintained that we experience "bodily modifications" unlocking "shames and indignations and fears" that are brought about by our sensitivities to another's perception of the self (James, 1884, p. 195). As ultra-social organisms who stand or fall by our social reputation, moral emotions motivate us to perform socially valued acts while affectively prohibiting socially disruptive ones (Keltner & Buswell, 1997; Tangney et al., 2007). One may therefore expect them to wield large physiological responses, corresponding with their strong motivational roles (Williams & DeSteno, 2008).

Guilt and pride are moral self-conscious emotions integral to preserving social bonds. Guilt is an intense, gnawing feeling of moral discomfort experienced when one's behaviour violates a personal or societal standard (Baumeister et al., 1994). Because guilt is also associated with empathy toward the victim, it can motivate reparative actions (Hoffman, 1998). Conversely, pride is a positive emotion that accompanies both our trivial and life-changing accomplishments (Tracy & Robins, 2004b). It therefore provides the psychological motivation or reinforcement for future pride-eliciting behaviours (Williams & DeSteno, 2008). The authentic rather than hubristic form of pride is considered adaptive in that it promotes continued achievement-oriented as well as prosocial behaviours (Tracy & Robins, 2007a). Both guilt and pride therefore motivate prosocial or socially adaptive behaviour, despite being associated with quite different phenomenological experiences.

Emotions initiate both motor responses and complex goal-directed behaviour in order to confer an adaptive advantage upon the individual (Cosmides & Tooby, 2000; Frijda, 1986). By this view, the primary function of physiological arousal may not be to determine the emotional response *per se*, but rather to prepare the body for different action programs recruited by the relevant emotion (Davidson, 1994). I therefore hypothesized that the physiology of guilt and pride should reflect their distinctive activation functions: Guilt should disrupt ongoing behaviour and operate as a punishment cue (Monteith et al., 2002), i.e., function as a source of negative arousal; pride, in contrast, should reinforce current behaviour and encourage one to perform well again (Tracy & Robins, 2007a), i.e., function as a source of positive arousal.

To understand physiological substrates underpinning behavioural motivations associated with guilt and pride more fully, I developed an emotion elicitation paradigm designed to be ecologically valid and intense enough to arouse the autonomic nervous system (ANS). I utilized measures of both sympathetic (SNS) and parasympathetic (PNS) nervous system activity to be able to discriminate between these modes of autonomic control during visceral arousal (Cacioppo et al., 2000).

Porges' (2001, 2007) polyvagal theory emphasizes how different physiological states support distinct forms of behaviour: Whereas cardiac SNS activation serves to mobilize bodily resources to cope with threats (i.e., fight or flight), cardiac PNS activation, via the vagus nerve, provides inhibitory input to the heart and facilitates social interaction (see also Porges, 1995). To index vagal activity, I analyzed the high frequency (HF; > .15Hz) and low frequency (LF; .04 - .15Hz) components of heart rate variability (HRV) by way of spectral analysis. While the HF variability in the heart is broadly accepted as a relatively pure estimate of vagal control (Berntson et al., 1997; Hayano et al., 1991), the physiological underpinnings of the LF component are more controversial (Goedhart, Willemsen, Houtveen, Boomsma, & de Geus, 2008). Conceptualizations of LF include it reflecting either a combination of vagal and sympathetic activity (Malliani et al., 1991; Martinmäki et al., 2006; Pagani et al., 1991), or baroreflex-mediated vagal activity (Brychta, Shiavi, Robertson, Biaggioni, & Diedrich, 2007; Moak et al., 2009). Despite the uncertainty regarding the source of LF variability in the heart rate, most researchers agree that reduced total HRV (both tonic and excessive reactivity) is nonadaptive and related to increased cardiovascular risk (Thayer & Lane, 2007).

The measures employed furthermore allowed me to distinguish between different modes of SNS arousal, given that distinctive regulatory mechanisms are thought to be involved in cardiac and skin sympathetic activity (Rainville et al., 2006): Whereas electrodermal activity is indicative of *somatic* arousal mediated predominantly by sympathetic cholinergic activity of human sweat glands (Dawson, Schell, & Filion, 2000), preejection period (PEP) is indicative of *cardiac* SNS arousal mediated by β-adrenergic inotropic drive (Cacioppo et al., 1994). PEP and SCL reactivity thus originate from largely distinct physiological systems that do not necessarily correlate with each other.

These SNS indices have been linked to the two primary motivational systems proposed by Gray, viz., the behaviour activation system (BAS) and the behaviour inhibition system (BIS) (Gray, 1982, 1987b; Gray & McNaughton, 2000). As mentioned previously, Beauchaine (2001) has suggested that both motivational systems are mediated peripherally by the sympathetic nervous system. Specifically, electrodermal activity has been associated with both trait and state anxiety (Katkin, 1965), which are mediated by the BIS (Fowles, 1980; Gray & McNaughton, 2000). In contrast, PEP has been linked to BAS activation because of the functional role of the SNS in mobilizing energy resources for behavioural activation (Beauchaine, 2001; Beauchaine et al., 2007; Brenner et al., 2005).

Because so much is still unknown about the duration of affective phenomena (Verduyn et al., 2009), and therefore the time course of any concomitant patterned autonomic response, there is considerable variability in the literature in terms of the temporal matching between the measurement of an emotion, and the onset of emotion induction (Levenson, 2003a). For example, while most studies have recorded physiological data *during* the emotion induction (Boiten, 1996; Britton, Taylor et al., 2006; Khalfa et al., 2002), some studies have collected physiological data *after* the emotion induction (Schwartz, Weinberger, & Singer, 1981; Stemmler, 1989). To ensure that I captured the emotional response in full, I recorded and analyzed physiological data both during the emotion manipulation, as well as directly afterward, during the post-emotion manipulation period. This approach has also been adopted in previous emotion research, and indicated that the effects of an emotion manipulation may even become more pronounced during the post-manipulation period (see, e.g., Herrald & Tomaka, 2002).

The experimental paradigm relied on interpersonal induction through staged interactions with two confederates to elicit real emotions of guilt and pride. Contemporary emotion research relies increasingly heavily on such methodology, which is believed to hold more merit and internal validity than hypothetical or remembered scenarios (Harmon-Jones et al., 2007; Herrald & Tomaka, 2002; Williams & DeSteno, 2008). Furthermore, it has been suggested that the emotion-triggering process may be dependent on, or enhanced by, the sociality of the stimulating context (Britton, Phan et al., 2006). The use of two carefully trained female confederates greatly enhanced the realism of the setup, so that participants actually experienced the emotions in question.

A pilot study that was conducted in the department validated the effectiveness and credibility of the interpersonal induction technique. One of the most promising findings emerging from the pilot study was that the elicitation of guilt, using this paradigm, appeared to be reasonably uncontaminated by the similar emotion of shame, as well as other negative emotions such as fear and anger. In the Guilt condition, participants were led to believe they were partly responsible for the dismissal of a research assistant. The manipulation was therefore designed to elicit guilt resulting from empathy for a victim, combined with the belief that one has had an active part to play in another's distress (Frijda, 1994; Hoffman, 1998). In addition, the scenario was one of morality, which was deemed appropriate because research suggests that guilt arises primarily in response to moral transgressions (Ferguson & Stegge, 1995). In short: at the outset of the experiment, a confederate acting as the research assistant offered each participant double the amount of money the participant was supposed to receive as study compensation. Participants were told not to mention this to anyone. Halfway through the experiment, however, a confederate acting as the supervisor discovered that money was missing from the participant compensation fund. Upon learning the truth of the missing money, the supervisor then decided to dismiss the research assistant, leaving participants feeling responsible and guilty for accepting the additional money, as well as for the ill fortune that befell the research assistant.

For the Pride condition, I employed preprogrammed positive feedback as well as staged interactions with confederates (similar to, e.g., Herrald & Tomaka, 2002; Williams & DeSteno, 2008), given that pride is most strongly evoked in situations of publicly praised accomplishment (Webster, Duvall, Gaines, & Smith, 2003). Pride in the current context resulted from participants' perceived accomplishments on an attentional task, and may be characterized as the authentic, more achievement-oriented form of pride that is thought to serve an adaptive social role (Tracy & Robins, 2007a). Most theorists agree that authentic pride can be defined as "a positive, self-conscious emotion arising from achievements that can be attributed to one's abilities and efforts" (Williams & DeSteno, 2008, p. 1007).

I also included a Neutral control condition to distinguish physiological responses specific to emotion induction from those due to the experimental context (Stemmler et al., 2001). In this condition, participants experienced the same context and sequence of events as in the other conditions, but without any emotional overlay. The Neutral condition thus served to control for any potentially unrelated sources of variability in the data (Christie & Friedman, 2004).

Based on previous research suggesting greater autonomic activation in negative than in positive emotions (the so-called "negativity bias"; Cacioppo et al., 2000; Cacioppo & Gardner, 1999), I predicted that guilt would be characterized by greater cardiac reactivity than pride. However, because no study to date has investigated, in a direct manner, the cardiovascular physiology of a current guilt experience, I had no specific hypothesis regarding ANS contributions in this condition. Because of the internal apprehension and anxiety caused by feelings of guilt, I predicted significant sympathetic arousal. While I also expected greater electrodermal activity in the guilt versus pride condition, I did not expect this measure to differentiate between these two emotion conditions, because of its correspondence to arousal rather than valence (Cacioppo et al., 2000). For pride I anticipated, based on previous findings (Herrald & Tomaka, 2002), low cardiovascular arousal.

A related aim of the present investigation was to assess individual difference characteristics in the experience of guilt and pride. Various self-report scales were therefore included to assess their relation to subjective as well as physiological emotional responses. Notably, all participants completed the self-report BIS/BAS scales (Carver & White, 1994) in order to determine the relation between BIS and BAS sensitivity and experimentally-induced affect. I predicted that, because of guilt's function as a punishment cue, that a high BIS score would be associated with greater experimentally-induced guilt, whereas higher sensitivity for personal reward (i.e., a high BAS score), would be associated with greater experimentally-induced pride (Heponiemi, Keltikangas-Järvinen, Puttonen, & Ravaja, 2003).

Method

Participants

Fifty-six female participants between the ages of 18 and 25 were recruited from a university population. I recruited only females to avoid confounds due to sex differences in emotional experience, expression of negative emotions, and physiological responses (Manstead, 1992; Shields, 1991; Tangney & Dearing, 2002). In addition, because both confederates were White, an all-White participant sample was recruited to avoid the potential confound of differing cross- and inter-racial attitudes from interfering with the desired emotional response in the social interaction task (Dovidio, Kawakami, & Gaertner, 2002).

Health status of participants was determined by self-report questionnaire (presented in Appendix A). Those with previously diagnosed neurological, psychiatric, cardiovascular, or substance use disorders were excluded. Other exclusion criteria included undergoing medical treatment affecting the circulation, taking any medication for depression or anxiety in the previous 6 months, and being left-handed (Oldfield, 1971). Contraceptive medication was allowed. Participants were informed that they would receive ZAR30 as compensation for participating in the 90-min study. The study protocol was approved by the Research Ethics Committee of the University of Cape Town's Department of Psychology.

All participants completed the procedures described below. Seven were excluded before data analysis, however, because of equipment failure (n = 2), high scores (> 26) on the Beck Depression Inventory II (BDI-II; Beck, Steer, & Brown, 1996) (n = 2), and post-experimental interviews suggesting they suspected deception (n = 3). The final sample thus consisted of 49 healthy females (age: M = 19.92 years, SD = 1.58).

Experimental Design and Setting

Each participant was randomly assigned to either a Guilt (n = 16), Pride (n = 16), or Neutral (n = 17) condition in a cross-sectional between-subjects design. Because the paradigm involved deception, a within-subjects design was not feasible.

The study took place within a dedicated research laboratory in the Department of Psychology at the University of Cape Town. A section of the lab was enclosed with curtains to allow for privacy of physiological recordings. Participants were, however, still able to overhear conversations in the rest of the laboratory. The curtained-off section housed two computers (one for data acquisition and one dummy), as well as the ambulatory monitoring system and electrodes used for physiological recordings.

Computerized Tasks

Computerized tasks were presented on a PC with a 13-in. monitor in three blocks, so that block consisted of three different tasks. The first task in a block was always a simple reaction time task that required very little attention. It consisted of a moving virtual environment of hospital corridors; participants simply had to press a key every time a doorway was passed through. The second task was a *Corsi Block*-type test where participants had to remember and repeat progressively longer spatial sequences. The final task in a block was a *Digit Span*-type task where participants had to remember and repeat a progressively longer number sequence. Both working memory tasks stopped after two consecutive errors.

Psychological and Self-Report Measures

Depression. The Beck Depression Inventory II (BDI-II; Beck et al., 1996) was employed to measure the presence and severity of depressive symptomatology. It is a 21-item multiple-choice self-report measure in which participants are asked to rate to what extent they have experienced each symptom during the past two weeks on a 4-point scale, ranging from 0 to 3. The questionnaire produces a single score as a measure of the intensity of the depressive symptoms, with higher scores indicating more cognitive, motivational, behavioural, and somatic symptoms of depression.

The BDI-II has been used regularly in South African research (e.g., Ward, Flisher, Zissis, Muller, & Lombard, 2001), and has good test-retest reliability (Beck et al., 1996). In addition, the BDI-II exhibits a high level of internal consistency in a student sample, with alphas ranging from .89 to .93 (Whisman, Perez, & Ramel, 2000).

Dispositional affect. General levels of positive and negative affect were assessed using the profile of Positive and Negative Affective Schedule (PANAS; Watson, Clark, & Tellegen, 1988). The PANAS is a self-report measure that consists of two mood scales with 10 items each (e.g., enthusiastic, distressed, jittery) to provide brief measures of positive and negative affect. Respondents are required to rate each particular emotion according to the strength of their experience on a 5-point scale, ranging from 1 (*very slightly or not at all*), to 5 (*extremely*). The scales can be used to detect current mood fluctuations when used with the short-term instructions (e.g., “right now” or “today”), or can be used to measure more general affective traits when the longer-term instructions are used (e.g., “past year”). Participants in the current study were asked to indicate how they feel “generally”; the measure was therefore used to determine pre-experiment levels of positive and negative affect.

The scales have been shown to have excellent convergent and discriminant validity and exhibit traitlike stability over a 2-month time period when the longer-term instructions are used (Watson et al., 1988). The PANAS is also reliable in several settings, and Cronbach’s alphas range between .85 and .89 for the PA scale, and between .82 and .85 for the NA scale (Crawford & Henry, 2004; Thompson, 2007).

Proneness to guilt, shame and pride. The Test of Self-Conscious Affect-3 (TOSCA-3; Tangney, Dearing, Wagner, & Gramzow, 2000) was used to assess participants’ dispositional Shame-Proneness, Guilt-Proneness, Alpha Pride and Beta Pride. The TOSCA-3 is a scenario-based measure that consists of 16 common scenarios drawn from written accounts of personal guilt, pride, and shame experiences of several hundred college students. Five of these scenarios have positive outcomes (e.g., “You and a group of co-workers worked very hard on a project. Your boss singles you out for a bonus because the project was such a success”), and the remaining 11 have negative outcomes (e.g., “At work, you wait until the last minute to plan a project, and it turns out badly”). For each scenario, participants are given a set of responses and are asked to indicate, on a 5-point scale, their likelihood of responding with shame, guilt, externalization, alpha pride, beta pride, and detachment/unconcern.

Construct validity has been demonstrated in the TOSCA’s relation to other measures of shame and guilt, as well as to other theoretically relevant constructs (Tangney, 1995; Tangney, Wagner, & Gramzow, 1992). Test-retest stability has been shown to be satisfactory (.85 and .74 for shame and guilt, respectively; Tangney, Wagner, Fletcher et al., 1992), while alphas for the TOSCA-3 subscales ranged between .46 and .74 in a recent study (Hasui et al., 2009).

Behavioural activation and inhibition. Tendencies toward behavioural inhibition and behavioural activation sensitivity were assessed through the BIS/BAS scales (Carver & White, 1994). The BIS/BAS scales consist of 20 items (plus 4 distracter items) that describe behavioural tendencies, e.g., “When I want something I usually go all-out to get it”, or

"Criticism or scolding hurts me quite a bit." Respondents are required to rate each statement on a 4-point Likert-type scale ranging from 1 (*very true for me*) to 4 (*very false for me*).

The BAS scale (13 items) is related to extraversion, positive affectivity, and increased sensitivity to cues of reward. It can be divided into three subscales that assess the self-reported tendency of an individual to (i) respond with drive toward incentives (BAS Drive), (ii) be highly responsive to reward in the environment (BAS Reward Responsiveness), and (iii) seek out fun activities (BAS Fun Seeking). Conversely, the BIS scale (7 items) measures apprehension or inhibition in response to a threatening situation and is associated with greater proneness to anxiety, neuroticism, and negative affectivity in general. High scores on the BIS scale are associated with increased sensitivity to cues of punishment or nonreward.

The BIS/BAS scales possess good test-retest reliability (Carver & White, 1994), and Cronbach's alphas for the BIS and BAS subscales have been shown to be 0.76 and 0.83, respectively (Jorm et al., 1999).

State affect. Measures of state affect were obtained at baseline (Emotion Time 1) and directly after emotion manipulation (Emotion Time 2). Although the second affect measure included the same items as the first, the order and visual presentation of items differed to disguise the fact that it was a repeated measure. Words describing different affective states were presented individually on the computer interface so that participants could rate their current emotional state on a 9-point Likert-type scale, ranging from 1 (*not at all*) to 9 (*very much*). Single-item measures included guilt, shame, pride, satisfaction, anger and anxiety. In addition, indices were created for general negative affect (averaged ratings for *sadness, fear* and *disgust with situation*) as well as a general positive affect (averaged ratings for *happiness, hope* and *relief*). This index creation was done to reduce the number of dependent variables for subsequent analyses. My primary interest, however, was in single-item ratings of guilt and pride.

Prosocial motivation. A questionnaire on HIV/AIDS was included to assess amending or prosocial behaviour after the emotion manipulation. The first section of this questionnaire assessed general knowledge about AIDS. Participants were then asked to indicate whether they would be willing to volunteer their time as a research participant in various HIV research studies. If they were willing, they had to indicate for how many studies, ranging from 1 to 6 (30 min each), they would volunteer over a period of 1 year. The questionnaire highlighted the fact, however, that they would not receive any financial compensation for their participation.

Manipulation check. Upon completion of the study, participants were shown a list of six possible emotions and asked to make a forced-choice decision as to which emotion, as well as the intensity level (on a Likert scale ranging from 1 'not at all', to 5 'very much'), they experienced most during the experimental manipulation. The list of emotions included fear, pride, shame, guilt, anxiety, and satisfaction; there was also a neutral option (if the participant felt no particular emotion). Participants were allowed to select a maximum of three choices.

Procedure

On arrival at the laboratory, each participant signed consent documents (Appendix B) and completed sociodemographic, mood, and personality measures (i.e., BDI-II, PANAS, BIS/BAS, TOSCA-3). The experimenter briefly explained the nature of the investigation and gave each participant a standard set of orienting remarks. The experiment's cover story was that participants would practice some computerized working memory tasks to assess associated physiological change (Appendix C). Each participant was then guided to the curtained-off section of the laboratory where the computerized tasks were explained to them. Two actors performed as confederates to heighten the realism of the setup. They were both female to avoid potential gender biases from interfering with the emotion induction (all participants were also female). The dialogue between the experimenter and her confederates was carefully scripted, as were the lines the experimenter and confederates delivered to the participants. (Appendix C presents the emotion manipulation dialogue.)

Pre-manipulation procedures. The experimenter explained to participants that a research assistant (Confederate #1) would assist in task administration because she (the experimenter) was needed elsewhere and could not stay for the full testing period. While the experimenter introduced the experimental procedures, the supervisor (Confederate #2) entered the laboratory and requested that the experimenter meet with her as soon as possible. The experimenter therefore left the laboratory while the assistant remained with the participant. The assistant attached all electrodes necessary for physiological measurements and once signal integrity was established, recorded a baseline rest period (5 min). Participants then rated their current emotional state (Emotion Time 1). The rationale for doing this was explained to participants in terms of affect changes or transient emotional states that might influence a person's cognitive skills. It was therefore necessary to periodically assess affect during the course of an experiment which would, ostensibly, allow the research team to control for small shifts in mood that might influence results (Carver & White, 1994). The research assistant also explained that no verbal communication was allowed during the experiment to ensure signal integrity.[2]

Experimental manipulation: Guilt condition. In this condition, the research assistant nonchalantly gave participants more money than the normal compensation amount (ZAR60 instead of ZAR30), indicating that this offer was not standard procedure and should be kept secret. As justification for the extra compensation, the assistant mentioned that some participants did not show up for testing and that the supervisor would not miss the money. The assistant left “to go to the bathroom” during the second block of computerized tasks. Shortly thereafter, the supervisor and experimenter unexpectedly returned. This entrance was timed to take place as participants finished the second block of computerized tasks. Upon discovering money missing, the supervisor interrogated the participant, asking her directly whether she had been given more than ZAR30 by the research assistant. When participants confirmed this through a nod of the head (they were still under instructions not to speak), the supervisor communicated her feelings that the research assistant was untrustworthy and slack and needed to be fired. The supervisor furthermore instructed the experimenter to hire a new assistant as soon as possible, and to finish off the experiment while she dealt with the assistant. Participants were therefore left to understand they had some agency in the misfortune of the kind assistant. Shortly thereafter, a second measure of state affect (Emotion Time 2) required participants to rate their current emotional state.

Experimental manipulation: Pride condition. Pride condition participants underwent a similar procedure as those in the Guilt condition, but with these exceptions: (i) They were not offered extra money, and (ii) they received bogus visual and verbal performance feedback aimed at eliciting achievement-oriented pride (e.g., Williams & DeSteno, 2008). During the instruction phase, the experimenter informed the participant that her test scores would be transferred to a ‘host-computer’ (i.e., the dummy computer). Participants in this condition also received predetermined normative information suggesting superior performance (e.g., a high percentile score and congratulatory sentence) after each computerized task block (Webster et al., 2003).

[2]This control measure was enforced in all three conditions to minimize physiological confounds due to respiratory changes.

When the supervisor and experimenter returned unexpectedly midway through the tasks, the supervisor ostensibly noticed the exceptional scores on the 'host computer' and warmly congratulated the participant. She used scripted nonverbal cues, such as smiling, voice intonation and shaking the participant's hand, as well as phrases such as "you are outperforming most other candidates" and "keep it up" to convey how impressed she was with the participant's performance.

Experimental manipulation: Neutral condition. Neutral condition participants experienced a similar context to those in the emotion elicitation conditions, but did so without any emotional overlay. They thus did not receive extra money or performance feedback, and all dialogue between confederates, as well as questions posed to the participant, featured neutral content. Care was taken to ensure that the amount of personal interaction in this condition matched that in the two emotion elicitation conditions. In the Neutral condition, the supervisor only inquired of the participant whether everything was progressing satisfactorily. This condition served to determine the magnitude and direction of physiological changes specific to emotion elicitation and distinct from the experimental context (Stemmler, 1989).

Post-manipulation procedures and debriefing interview. Upon completion of the third block of computerized tasks, participants completed the HIV/AIDS questionnaire to assess prosocial activation. Participants were then told that the experiment was over and that they were allowed to talk again. They were then asked to indicate which emotion they experienced most when the supervisor entered for the second time (i.e., the manipulation check). Importantly, the manipulation check was performed *before* the experimenter overtly stated the real purpose of the investigation. Where necessary, the experimenter clarified certain emotion terms (e.g., shame and guilt) to avoid artifacts due to erroneous labeling of emotions.

The debriefing session took the form of a funneled approach, where the experimenter first carefully probed participants for suspicion while maintaining the cover story, followed by more specific questions relating to their experiences during the experiment (Harmon-Jones et al., 2007). To distinguish between guilt and shame, participants' counterfactual thinking was also recorded by asking them to imagine what might have caused the event to end differently. For example, they were asked: "If you could change what happened, do you wish that you were a different type of person, or do you wish that you responded differently?" (Niedenthal et al., 1994). The debriefing session thus served an important purpose in verifying that target emotions were actually experienced (Levenson, 2003a). Participants were then debriefed in full about the true nature of the investigation and were asked not to discuss the experiment with fellow students. In particular, participants in the Guilt condition were reassured that the research assistant had not been fired, and they were allowed to keep the extra money as compensation for the distress they had experienced.

Physiological Measures

Ambulatory Recording. The Vrije Universiteit Ambulatory Monitoring System (VU-AMS, Version 5fs; de Geus & van Doornen, 1996) recorded the electrocardiogram (ECG), impedance cardiogram (ICG), and skin conductance level (SCL) continuously during the experiment. Participants remained seated throughout the procedure and were instructed to refrain from talking and making exaggerated body movements. These precautions were taken to ensure that varying respiratory behaviour did not affect physiological recordings (Grossman, Wilhelm, & Spoerle, 2004).

During the experiment, event markers were inserted at predetermined time-points in the manipulation to designate: (i) A neutral period of 90 s that coincided with a button-press task of low attentional load; (ii) the emotion manipulation (EM) period of 90 ± 5 s, starting with the entrance of the supervisor and experimenter halfway through the experiment and ending with the supervisor's exit; and (iii) the post-emotion manipulation (Post-EM) period that consisted of the 90 s directly after the emotion manipulation and also coincided with the low-load attentional task. A baseline period of 5 min was also recorded. All periods were kept of identical length across different experimental conditions. Because of the nature of the emotion manipulation (i.e., because it depended on the delivery of scripted sentences), the duration of this period varied slightly across participants. Event markers were used as guides to extract physiological data from fixed periods of interest, i.e., from the neutral, emotion manipulation, and post-emotion manipulation periods.

Ambulatory Signal Scoring. The ambulatory monitoring procedure has been described in detail elsewhere (Goedhart, Kupper, Willemsen, Boomsma, & de Geus, 2006; Goedhart, van der Sluis, Houtveen, Willemsen, & de Geus, 2007; Riese et al., 2003). Briefly, the ECG was recorded from three disposable, pregelled Ag-AgCl electrodes attached in a triangular, equidistant configuration on the precardium, with signals sampled at 500Hz. The recorded interbeat interval (IBI) time series was visually inspected for physiologically implausible readings. Artifacts were corrected by summing spuriously short IBIs, while missing beats were 'created' by splitting spuriously long IBIs (> 99% of data were free of artifacts).

From the ECG tachograms, I calculated HR as well as measures of heart rate variability (HRV). HRV was estimated through the time domain measure RMSSD, i.e., the root mean of the squared successive differences of successive heartbeat intervals, as well as power spectral densities via autoregressive (AR) analysis. RMSSD has been found to correlate strongly with the high frequency (HF) component of frequency-domain methods, and vagal activity in turn is the major physiological contributor to the HF component (Berntson, Lozano, & Chen, 2005; Task Force, 1996). RMSSD could therefore be viewed as reflecting largely parasympathetic modulation of the heart through vagal activation/withdrawal, although lower spectral frequencies also influence this metric. The AR analyses were done via a software package from the Biomedical Signal Analysis Group (Department of Applied Physics, University of Kuopio, Kuopio, Finland), using a model order of 15 and an interpolation rate of 5 Hz. To obtain valid HRV estimates, I extended the neutral and emotion manipulation periods for these analyses from 90 s to 120 s (Task Force, 1996). Frequency bands assessed included respiratory or high frequency (HF; .15 - .40Hz), low frequency (LF; .04 - .15Hz), and the sum of HF and LF, total frequency (TF; .04 - .4Hz). All calculated $msec^2$ power values were transformed via natural logarithm to normalize the distributions.

The ICG was monitored using a four spot-electrode configuration consisting of two electrodes at the back, which supplied high-frequency current, and two measuring electrodes on the chest, to detect the voltage drop over the thorax. The electrical resistance through the chest was thus measured as a function of blood volume variation while passing a constant current of 350 µA, 50 kHz through the chest cavity. Resulting measures included basal thoracic impedance (Z_0) and the first derivative of basal impedance (dZ/dt, sampled at 500 Hz), from which several systolic time interval indices of cardiac contractility (e.g., PEP) could be calculated. The impedance data were ensemble-averaged over 30-s intervals, and a manual scoring procedure detailed by Sherwood et al. (1990) was used to identify certain waveform components necessary to calculate PEP. PEP scoring was quantified as the time interval in milliseconds between the ECG Q-wave and the B-point in the ICG, which is the start of the rapid upslope of the dZ/dt waveform to its maximum. PEP was calculated as the mean value of all 30-s ensemble averages within fixed periods of interest (Riese et al., 2003).

SCL was recorded as an index of the magnitude of emotional arousal, independent of valence (Lang, Greenwald, Bradley, & Hamm, 1993). The constant voltage method (0.5 V), sampled at 10Hz, was used to measure changes in electrodermal activity in standard microSiemens (μS) conductance units. Ag-AgCl, non-polarizable finger electrodes (6mm diameter contact area; BiopacSystems, Inc.) filled with isotonic, 0.5% saline gel (GEL101, BiopacSystems, Inc.) were attached to the distal phalanx surfaces of participants' middle and index fingers of the nondominant left hand. Participants were requested to rest their left hands on the table during the experimental procedure.

Statistical Analysis

Subjective response data from Emotion Time 1 and Time 2 were analyzed to determine whether the emotion manipulations were effective. Within-subjects analyses, via paired-samples *t*-tests, were performed to determine the relative increases/decreases in target emotions of participants in specific experimental conditions. At the group level (i.e., between-subjects), I performed multivariate analyses of variance (MANOVAs), with emotion condition as the between-subjects factor and ratings of affective states as the dependent variables. To determine which self-reported emotions contributed most to group separation, I performed a discriminant function analysis.

As noted above, all participants engaged in a 5-min baseline period and a 90-s neutral period that were identical across all three experimental conditions. Because the neutral period corresponded better to the emotion manipulation period in length and attention-demand, this period was used in all analyses as the control period against which to assess physiological change. Intercorrelations between physiological reactivity indices were performed to investigate relations between physiological variables. To determine emotion-specific physiological changes during the emotion manipulation (EM) and post-emotion manipulation (Post-EM) periods, one-way analyses of co-variance (ANCOVAs) were performed with emotion condition as the between-subjects factor and reactivity during the neutral period as the covariate.

Finally, individual difference analyses were performed by correlating personality measures with experimentally induced-affect, as well as with physiological response data.

Results

Results for Study 1 are presented in three parts: (i) Subjective Responses, (ii) Physiological Responses, and (iii) Individual Differences. There were no significant between-group differences in terms of age at testing ($p = .66$), years of education ($p = .46$), or scores on the Beck Depression Inventory II ($p = .12$).

Subjective Responses

Dispositional positive and negative affect. Analyses of general positive and negative affect, as assessed through the PANAS, revealed no significant between-group differences in dispositional positive affect (PA), $F(2,46) = 0.33$, $p = .72$, or negative affect (NA), $F(2,46) = 0.35$, $p = .71$.

State affect. The dependent variables were participants' self-reported levels of emotion before (Emotion Time 1) and after (Emotion Time 2) the experimental manipulation. If the guilt and pride manipulations were successful, participants in these conditions should report heightened levels of guilt and pride, respectively, relative to participants in the Neutral condition. Table 1 presents affective ratings of participants before and after the emotion manipulation.

Within-subjects analyses of state affect changes from pre-manipulation (Emotion Time 1) to post-manipulation (Emotion Time 2), showed that Guilt condition participants significantly increased their ratings of guilt, as well as those of anxiety, anger, general negative affect, and shame, $ts(15) > 3.50$, $ps < .01$, $rs > .60$. Their ratings of pride decreased slightly, $t(15) = 1.78$, $p = .09$, $r = .42$. Although ratings of guilt increased the most, $t(15) = 4.09$, $p = .001$, $r = .73$, this change was not statistically significant compared to increases in other negative emotions ($ps > .10$, $rs < .40$).

Pride condition participants significantly increased their ratings of pride, $t(15) = 3.26$, $p = .005$, $r = .64$, as well as those of general positive affect, $t(15) = 2.26$, $p = .04$, $r = .50$, but not satisfaction, $t(15) = 0.44$, $p = .67$, $r = .11$. Their ratings of guilt showed a significant decrease, $t(15) = 2.28$, $p = .04$, $r = .51$.

Neutral condition participants did not change their ratings of either guilt or pride from pre- to post-manipulation ($ps > .15$, $rs < .35$). Neutral condition participants, in fact, showed no significant changes on any emotion, besides a significant decrease in satisfaction, $t(15) = 2.52$, $p = .02$, $r = .53$.

Table 1

Changes in State Affect During the Three Experimental Conditions

		Neutral (n = 17)		Pride (n = 16)		Guilt (n = 16)	
		Mean	*SD*	Mean	*SD*	Mean	*SD*
Satisfaction:	Emotion Time 1	6.18	1.51	6.31	1.08	6.44	1.31
	Emotion Time 2	4.88	2.09	6.50	1.67	3.56	2.03
Anxiety:	Emotion Time 1	3.53	1.94	2.88	1.41	3.50	1.90
	Emotion Time 2	2.88	2.00	2.88	1.82	6.81	1.38
Anger:	Emotion Time 1	1.41	1.46	1.44	.96	1.38	.62
	Emotion Time 2	1.24	.44	1.44	.96	4.50	2.16
Pride:	Emotion Time 1	4.71	2.11	4.75	2.08	4.50	2.16
	Emotion Time 2	4.53	2.27	6.38	1.93	3.50	2.03
Guilt:	Emotion Time 1	1.65	1.50	2.44	1.75	3.31	2.33
	Emotion Time 2	1.53	1.46	1.63	.89	6.81	1.83
Shame:	Emotion Time 1	1.59	1.50	1.81	1.33	2.75	2.05
	Emotion Time 2	2.00	1.77	1.38	.72	5.38	2.22
General Neg:	Emotion Time 1	1.92	1.27	1.81	.89	2.58	1.41
	Emotion Time 2	1.87	1.29	1.63	.88	5.15	1.49
Sadness:	Emotion Time 1	2.00	1.70	1.94	1.29	2.69	1.89
	Emotion Time 2	1.88	1.41	1.94	1.48	4.94	1.95
Disgust with	Emotion Time 1	1.47	1.70	1.50	1.10	1.75	1.24
situation:	Emotion Time 2	1.65	1.17	1.25	.58	6.00	2.19
Fear:	Emotion Time 1	2.12	1.50	2.00	1.32	3.31	2.02
	Emotion Time 2	1.94	1.60	1.69	1.14	4.50	1.90
General Pos:	Emotion Time 1	4.94	1.35	5.25	1.02	5.60	1.23
	Emotion Time 2	4.60	1.54	5.90	1.34	4.29	1.72
Relief:	Emotion Time 1	3.06	1.68	4.37	1.59	4.31	1.78
	Emotion Time 2	3.88	1.62	5.44	1.86	3.94	2.35
Hope:	Emotion Time 1	5.35	1.87	5.44	1.03	5.69	1.92
	Emotion Time 2	4.94	2.16	5.94	1.48	5.38	2.42
Happiness:	Emotion Time 1	6.41	1.50	5.94	1.24	6.81	1.11
	Emotion Time 2	5.35	1.58	6.31	1.49	3.56	1.67

Note. Participants rated emotions along a 9-point Likert-type scale ranging from 1 (not at all) to 9 (very much). General Neg = general negative affect. General Pos = general positive affect.

To investigate affect differences across experimental groups, subjective ratings of all reported emotions were compared using multivariate analysis of variance (MANOVA).[3] The MANOVA conducted on the Emotion Time 1 data, with elicitation condition as the between-subjects factor and ratings of 8 affective states as dependent variables, was not statistically significant, $F(16,78) = 1.20$, $p = .30$, $\eta^2 = .19$ (Pillai's trace). All univariate ANOVAs for baseline emotions were also non-significant (ps > .10, rs < .30). There were thus no between-group differences with regard to baseline state affect. A similar MANOVA conducted on the Emotion Time 2 data, however, revealed significant between-group differences after the emotion manipulation, $F(16,78) = 5.28$, $p < .001$, $\eta^2 = .52$ (Pillai's trace).[4] All univariate ANOVAs were now significant (negative emotions: ps < .001; positive emotions: ps < .05), as illustrated in Figure 1.

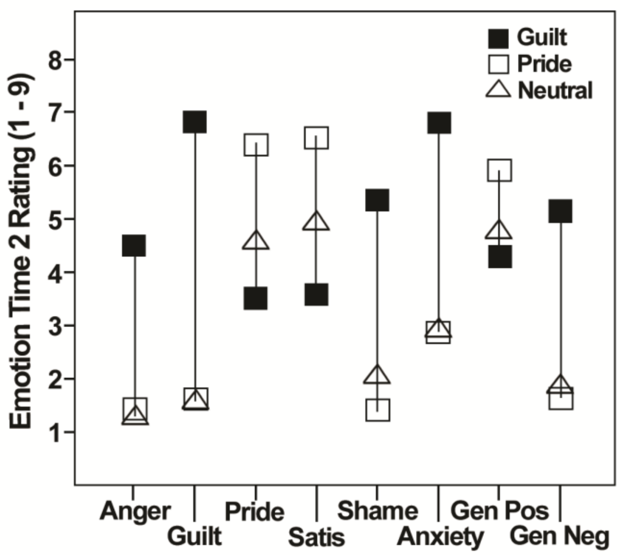


Figure 1. Graph illustrating differences between Guilt, Pride, and Neutral participants' mean state affect ratings after the emotion manipulation (i.e., Emotion Time 2 ratings).
Satis: satisfaction, Gen Pos: general positive affect, Gen Neg: general negative affect.

---

[3]Data were ln-transformed and groups equalized by excluding one randomly-chosen Neutral case to correct for any violations of Box's test.

[4]Because Box's test was significant at P < 0.01, the α-level was reduced from 5% to 2.5%. Pillai's trace, however, is assumed to be robust against violations of Box's test (Olson, 1976).

To determine which self-reported emotions contributed most to discriminating elicitation conditions from one another, I performed discriminant function analysis (DFA) on the Emotion Time 2 data. Although DFA does not always replicate well across studies, it was suitable in the current context because its purpose was mainly to establish whether the emotion manipulations were effective in eliciting guilt and pride, respectively. The DFA analysis yielded two discriminant functions (DFs): DF1 significantly discriminated the Guilt condition ($p < .001$), while DF2 discriminated the Pride condition, but not significantly ($p = .16$) (Figure 2). Structure coefficients indicated that guilt ($r = .85$), followed by general negative affect ($r = .61$), anger ($r = .59$), and shame ($r = .52$) contributed most to DF1 and therefore discrimination of the Guilt condition (Table 2). In contrast, pride ($r = .83$) and satisfaction ($r = .70$) contributed most to DF2 and therefore discrimination of the Pride condition (Table 2). Group classification rates for this analysis reached an overall accuracy of 85.4% (Guilt: 100%; Pride: 87.5% ; Neutral: 68.8%) (Table 3).

Taken together, the results confirmed that guilt and pride, respectively, were elicited more strongly than any other emotion by the emotion manipulation conditions.

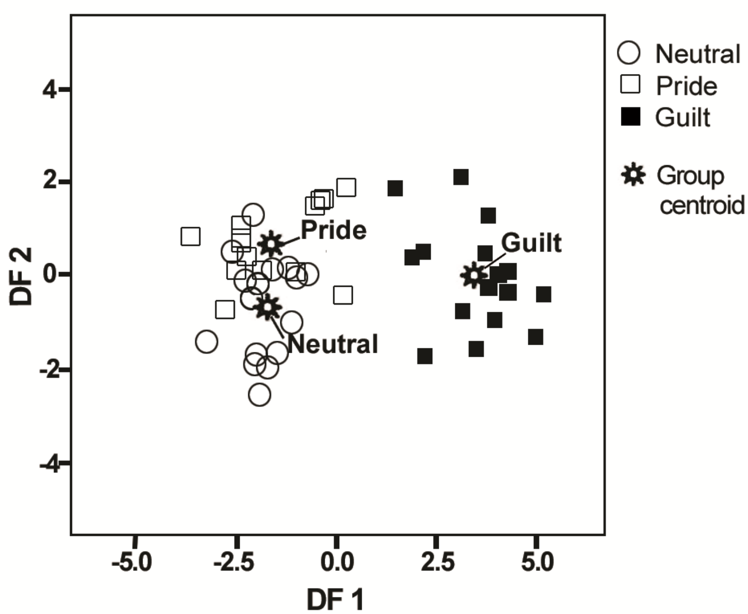


Figure 2. Discriminant function analysis plot of group centroids using self-report variables from Emotion Time 2 (i.e., after the emotion manipulation). The graph plots the variate scores for each participant, grouped according to the appropriate experimental condition.

Table 2

Structure Coefficients for the First and Second Discriminant Functions (DFs) of Emotion Time 2

	DF1	DF2
Guilt	**.85**	-.02
General negative affect	.61	-.29
Anger	.59	-.02
Shame	.52	-.45
Anxiety	.47	-.05
Pride	-.21	**.83**
Satisfaction	-.24	**.70**
General positive affect	-.14	.61

Note. Structure coefficients ≥ .70 are in boldface.

Table 3

Pattern Classification Accuracy of Emotion Conditions Based on Self-report Emotions from Emotion Time 2

	Predicted Group Membership			
Study Group	Neutral Count (%)	Pride Count (%)	Guilt Count (%)	Total
Neutral	11 (68.80)	5 (31.30)	-	16 (100)
Pride	2 (12.50)	14 (87.50)	-	16 (100)
Guilt	-	-	16 (100)	16 (100)

To explore subjective guilt and pride responses further, zero-order correlations were performed on Emotion Time 2 data for each experimental manipulation. In the Guilt condition, ratings of guilt were most significantly correlated with ratings of anxiety ($r = .83$, $p < .001$), followed by general negative affect ($r = .68$, $p = .004$), shame ($r = .64$, $p = .007$) and anger ($r = .55$, $p = .03$). Of the three emotions grouped together under 'general negative affect', guilt was only significantly correlated with disgust with situation ($p < .01$), but not with fear or sadness (*p*s > .11). Interestingly, guilt and pride had a significant negative linear relationship in the Guilt condition ($r = -.67$, $p = .004$). Feelings of pride therefore decreased as feelings of guilt increased.

In the Pride condition, satisfaction was the only emotion that correlated significantly with experimentally-induced pride ($r = .64$, $p = .008$).

Manipulation check analysis. The manipulation check was employed to verify that target emotions were successfully elicited during the experimental manipulations. Participants were required to choose, from a list of six, which emotions and with what intensity (from 1 to 5), they experienced most during the experimental manipulation. Within each experimental group, emotion ratings were averaged to obtain a rating out of 5.

Guilt condition participants rated guilt as the emotion they experienced predominantly ($M = 4.06$, $SD = 0.85$), followed by anxiety ($M = 2.88$, $SD = 0.87$), and shame ($M = 1.5$, $SD = 0.79$) (Figure 3). In the Pride condition, pride was the most highly-rated emotion ($M = 3.75$, $SD = 0.93$), followed by satisfaction ($M = 2.50$, $SD = 0.47$), and anxiety. Neutral condition participants primarily rated their mood as neutral throughout the experiment, although some also reported low levels of anxiety and satisfaction. The experimental context itself therefore evoked some anxiety.

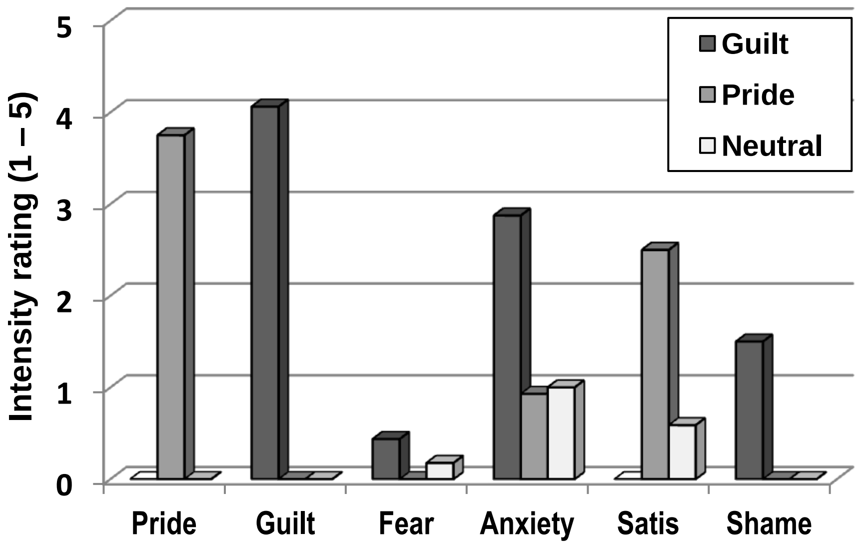


Figure 3. Post-experimental manipulation check ratings. Participants from the Guilt, Pride and Neutral conditions indicated which emotions they experienced most during the experimental manipulation. Satis: satisfaction.

Prosocial motivation. To assess prosocial motivation post emotion elicitation, the HIV/AIDS volunteer study data were analyzed. While there were no statistically significant between-group differences in the number of participants who volunteered to participate in further HIV/AIDS studies, $\chi^2(2) = 2.14$, $p = .34$, Pride condition participants were more likely to volunteer their efforts (68.75% volunteered), compared to participants in both the Neutral and Guilt conditions (47.01% and 43.75%, respectively). Pride condition participants on average also volunteered to participate in more studies per participant ($M = 2.25$, $SD = 2.28$), compared to participants in the Guilt ($M = 1.63$, $SD = 2.19$) and Neutral ($M = 1.29$, $SD = 1.72$) conditions, although these differences were not statistically significant ($p = .41$). Overall, Pride condition participants were thus more likely to engage in prosocial activities following the emotion induction.

Integrating the results reported so far, I conclude that the emotion elicitation procedures reliably and effectively produced the desired target affects of guilt and pride in the respective emotion manipulation conditions. The Neutral condition was also sufficiently neutral to be used as a control condition in subsequent physiological data analysis.

Physiological Responses

One-way analyses of variance (ANOVAs) detected no significant between-group differences on any physiological parameter during the neutral period (ps $> .20$, rs $< .25$).

Intercorrelations between physiological reactivity indices. Because no between-group differences existed during the neutral period, reactivity (Δ) scores could be calculated by subtracting the average value of each participant's neutral period from her emotion manipulation (EM) or post-emotion manipulation (Post-EM) period average. (For SCL, the neutral period mean was subtracted from the EM maximum value.) In this way, *additional reactivity* scores, which were associated with the emotional experience itself, were obtained for each physiological variable (Herrald & Tomaka, 2002). Figure 5 (see page 75) displays the average within-subjects additional reactivity or change scores during the emotion manipulation and post-manipulation periods. Spearman's correlation coefficients were computed to examine the relationships between reactivity scores of physiological variables in each experimental group. Because power spectral densities were calculated over slightly longer periods (120 s), these measures were analyzed separately.

Correlation results are shown in Table 4. Of particular interest are correlations between PEP and SCL reactivity scores. Although PEP and SCL both reflect sympathetic arousal, no significant correlations were observed between ΔPEP and ΔSCL for any experimental condition during the emotion manipulation (rs $< .35$, ps $> .18$). Another interesting finding was the dissociation in the heart rate response for the two emotion manipulation conditions: In the Pride condition, ΔHR during EM was significantly correlated with ΔRMSSD ($r = -.52$, $p = .04$), but not with ΔPEP; in the Guilt condition, however, ΔHR during Post-EM was significantly correlated with ΔRMSSD ($r = -.79$, $p < .001$) as well as ΔPEP ($r = -.76$, $p = .001$). Preliminary interpretation of these results suggest that heart rate increases in the Pride condition were predominantly due to vagal unloading, whereas heart rate increases in the guilt condition were a function of both vagal unloading and sympathetic drive. Consistent with previous reports (Cacioppo et al., 1994), changes in RMSSD and PEP reactivity were not correlated in any experimental condition.

Table 4

Spearman's Correlations Coefficients Among Physiological Reactivity Scores for Each Experimental Group

	1	2	3	4	5	6
Neutral Condition						
1. ΔSCL						
2. ΔHR EM	.33					
3. ΔHR Post EM	-	.18				
4. ΔRMSSD EM	-.06	-.36	-			
5. ΔRMSSD Post EM	-	-	-.45	.32		
6. ΔPEP EM	.31	.43	-	.02	-	
7. ΔPEP Post EM	-	-	.02	-	-.07	.88**
Pride Condition						
1. ΔSCL						
2. ΔHR EM	.14					
3. ΔHR Post EM	-	.47				
4. ΔRMSSD EM	-.01	-.52*	-			
5. ΔRMSSD Post EM	-	-	.29	.06		
6. ΔPEP EM	-.33	-.46	-	.09	-	
7. ΔPEP Post EM	-	-	-.47	-	.06	.89**
Guilt Condition						
1. ΔSCL						
2. ΔHR EM	.42					
3. ΔHR Post EM	-	.60*				
4. ΔRMSSD EM	.22	-.24	-			
5. ΔRMSSD Post EM	-	-	-.79**	.43		
6. ΔPEP EM	-.34	-.45	-	-.18	-	
7. ΔPEP Post EM	-	-	-.76**	-	.49	.79**

Note. SCL = skin conductance level; HR = heart rate; RMSSD = root mean of the squared successive differences; PEP = preejection period; EM = emotion manipulation period; Post-EM = post-emotion manipulation period.
**p* < .05. ***p* < .01.

Frequency analysis correlations. To explore relations among power spectral densities during the neutral and emotion manipulation periods, I performed zero-order correlations. As can be seen in Table 5, power spectral densities were significantly correlated *across* experimental conditions (rs > .45, ps ≤ .001). In particular, significant correlations between LF- and HF power are consistent with previous findings (Martinmäki et al., 2006), and indicative of significant vagal modulation in the low frequency component. Significant correlations between LF and HF power were also detected *within* each experimental group, with most correlations reaching statistical significance (.42 < *r*s < .72, .09 ≥ *ps* ≥ .001).

Table 5 also presents relations between the spectral metrics (i.e., LF, HF, and TF) and RMSSD, which can be seen to be significantly correlated.

Table 5

Zero-order Correlations Among Heart Rate Variability Measures Across Experimental Groups

	1	2	3	5	6	7
Neutral Period						
1. LF						
2. HF	.64***					
3. TF	.94***	.85***				
4. RMSSD	.70***	.87***	.87***			
Emotion Manipulation Period						
5. LF						
6. HF				.45***		
7. TF				.94***	.70***	
8. RMSSD				.68***	.77***	.81***

Note. Both the neutral and emotion manipulation periods were extended to 120 s for the frequency analyses, while the RMSSD measures were taken from the original 90-s periods. LF = low frequency power; HF = high frequency power; TF = total frequency power; RMSSD = root mean of the squared successive differences.
****p* ≤ .001.

Emotion manipulation effects.

SCL. The ANCOVA for SCL, with maximum SCL during EM as the dependent variable and mean reactivity during the neutral period as the covariate, was significant, $F(2,45) = 6.68$, $p = .003$, $\eta^2 = .23$. Post-hoc contrasts indicated significantly higher SCL for Guilt and Pride participants compared to Neutral participants ($p = .001$, $r = .47$ and $p = .03$, $r = .31$). SCL of Guilt and Pride participants did not significantly differ ($p = .20$), an indication that they experienced comparable levels of arousal relative to participants in the Neutral condition (see Figure 5A).

Because it has previously been reported that negative events have a stronger gain function than positive events in terms of producing affective reactions (Larsen, 2002), I tested this hypothesis by investigating the SCL decay magnitude of the respective experimental conditions. SCL decay was calculated by summing the effective area under the SCL graph for each participant from the point of maximum SCL (in the EM period) to precisely 90 s thereafter (Figure 4A).

Because SCL decay data showed large variability, a Kruskal-Wallis test was performed, with SCL decay as the dependent variable and experimental condition as the between-subjects factor. This analysis detected significant between-group differences, $H(2) = 7.36$, $p = .03$. Post hoc Mann-Whitney tests (using a Bonferroni correction and significance level of .025), indicated that the SCL response of participants in the Guilt condition endured for significantly longer than that of participants in the Neutral condition, $U = 67$, $p = .01$, $r = .43$. SCL decay of participants in the Pride condition also endured significantly longer than for participants in the Neutral condition, but only at the 5% level of significance, $U = 79$, $p = .04$, $r = .36$. A Joncheere-Terpstra test detected a significant trend in the data, namely that the SCL response was larger in magnitude (i.e., slower decay) as one moved from the Neutral to Pride to Guilt conditions: J = 544, $z = 2.64$, $r = .38$ (Figure 4B).

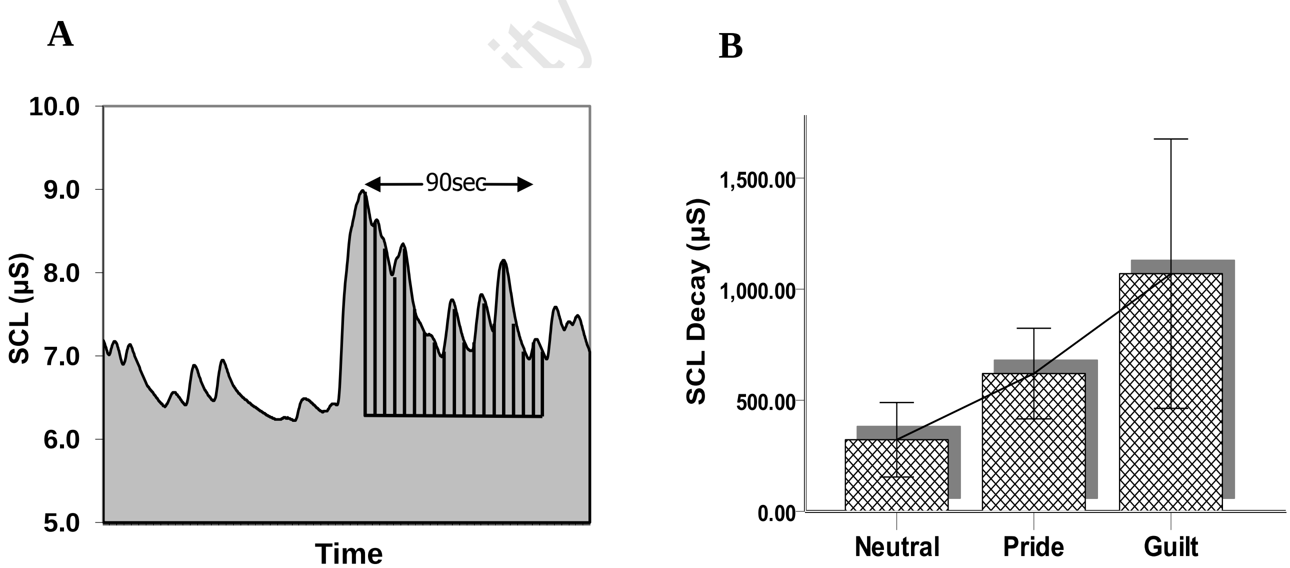


Figure 4. Skin conductance decay from SCLmax during the emotion manipulation period over 90 s. (A) Graphic illustration of the calculation of skin conductance decay. (B) Group differences in SCL decay magnitude.

HR. Because the HR distributions were skewed, HR data were reciprocally transformed to minimize large variability across participants. After excluding one outlier from the Guilt condition for both EM and Post-EM periods (more than 2 *SD*s from the mean), HR data complied with all assumptions underlying parametric data analysis.

The ANCOVA for HR detected significant between-group differences during EM, $F(2,44) = 6.20$, $p = .004$, $\eta^2 = .22$. Post-hoc contrasts indicated that HR of Guilt participants was significantly higher than that of Neutral participants ($p = .001$, $r = .43$), but not Pride participants ($p = .22$). HR of Pride participants was not significantly higher than that of Neutral participants ($p = .06$). Group differences in HR became more pronounced during Post-EM, $F(2,44) = 8.92$, $p = .001$, $\eta^2 = .29$, with Guilt participants' HR significantly higher than Neutral ($p < .001$, $r = .50$), as well as Pride ($p = .006$, $r = .38$), participants. Pride participants' HR still did not differ significantly from that of Neutral participants ($p = .23$) (Figure 5B).

PEP. PEP data were reversed-scored and ln-transformed to correct for negative kurtosis. After performing these transformations, PEP data complied with all assumptions underlying parametric data analysis.

The ANCOVA for PEP detected significant between-group differences during EM, $F(2,45) = 9.53$, $p < .001$, $\eta^2 = .29$. Similar to results for HR, these differences became more pronounced during Post-EM, $F(2,45) = 15.26$, $p < .001$, $\eta^2 = .41$. During both periods, significantly shorter preejection periods were observed for Guilt participants compared to Pride as well as Neutral participants (*p*s < .001, *r*s > .45). PEP of Pride participants did not differ from that of Neutral participants during either EM or Post-EM ($p > .40$) (Figure 5C).

RMSSD. For RMSSD analyses, two outliers were removed from the Neutral condition (these participants had exceptionally high RMSSD values, more than 2.5 *SD*s from the mean), after which data complied with assumptions underlying parametric data analysis.

The ANCOVA for RMSSD detected significant between-group differences during EM, $F(2,43) = 15.80$, $p < .001$, $\eta^2 = .42$, as well as during Post-EM, $F(2,43) = 7.27$, $p = .002$, $\eta^2 = .25$. During EM, the RMSSD of both Guilt and Pride participants were significantly lower than that of Neutral participants ($p < .001$, $r = .64$ and $p = .003$, $r = .42$). RMSSD of Guilt participants was also significantly lower than that of Pride participants ($p = .01$, $r = .37$). During Post-EM, RMSSD of Pride participants recovered somewhat and was not significantly different from Neutral participants anymore ($p = .20$). RMSSD of Guilt participants, however, remained lower than that of Neutral ($p = .001$, $r = .49$) as well as Pride ($p = .02$, $r = .36$) participants (Figure 5D).

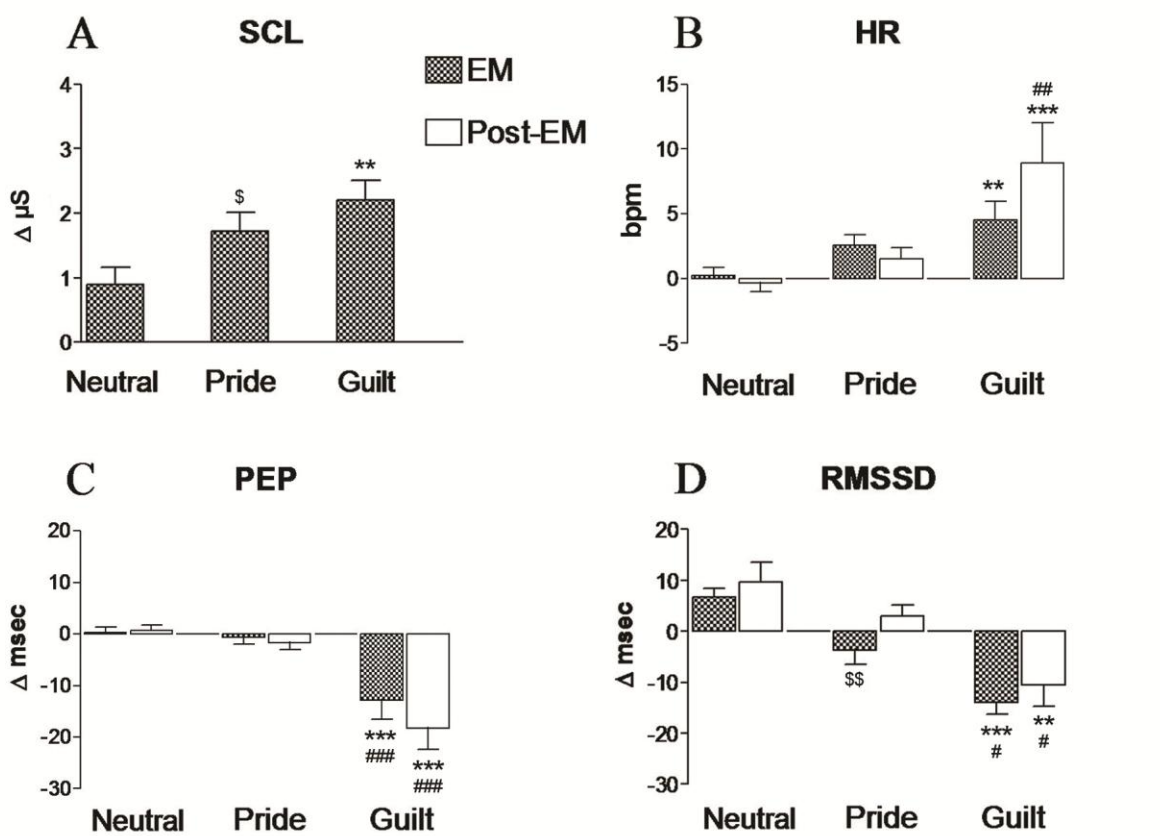


Figure 5. Physiological change scores (from the neutral period) for the Neutral, Pride, and Guilt conditions during EM and Post-EM. (A) SCL, (B) HR, (C) PEP and (D) RMSSD
EM: emotion manipulation period, Post-EM: post-emotion manipulation period, SCL: skin conductance level, HR: heart rate, PEP: preejection period, RMSSD: root mean of the squared successive differences.
$^{**}p < .01$, and $^{***}p < .001$ Guilt compared to the Neutral condition. $^{\$}p < .05$ and $^{\$\$}p < .01$ Pride compared to the Neutral condition. $^{\#}p < .05$, $^{\#\#}p < .01$, and $^{\#\#\#}p < .001$ Guilt compared to the Pride condition.

HR analysis with ΔPEP and ΔRMSSD as covariates. Because HR is dually controlled by the parasympathetic and sympathetic nervous system, I conducted ancillary analyses of the HR effects to see whether it was dependent on PNS, SNS, or both. Toward this aim, ΔPEP and ΔRMSSD were added as additional covariates in the HR ANCOVAs for EM and Post-EM. Because ΔPEP and ΔRMSSD were not significantly correlated in any of the experimental conditions, it is reasonable to assume that they constituted different aspects of ANS activity. If including PEP reactivity as a covariate eliminated or reduced the effects of the emotion manipulation on HR, this would suggest that SNS activity underlies HR increases during the emotion manipulation. By comparison, if RMSSD reactivity as a covariate influenced the effects of the emotion manipulation on HR, this would suggest that PNS activity underlies HR increases. ΔPEP and ΔRMSSD were added separately as covariates in the analyses to assess their individual effects in reducing the unexplained variance.

Results supported a pattern of increased sympathetic influence coupled with parasympathetic withdrawal for guilt. The ANCOVA with HR during EM as the dependent variable and HR during the Neutral period as well as ΔPEP as covariates, was not significant, $F(2,43) = 2.90$, $p = .07$, $\eta^2 = .12$. Similarly, the ANCOVA for HR during EM with HR during the Neutral period and ΔRMSSD as covariates, was not significant, $F(2,42) = 1.20$, $p = .31$, $\eta^2 = .05$. In these analyses, both ΔPEP and ΔRMSSD covariates were significantly related to HR increases during the emotion manipulation period ($p = .02$ and $p = .01$, respectively). Controlling separately for the effects of ΔPEP and ΔRMSSD on HR thus led to non-significant changes across experimental groups.

Similar results were obtained for HR when controlling for ΔPEP and ΔRMSSD during the post-emotion manipulation period. Adding ΔPEP and ΔRMSSD separately as covariates in this analysis resulted in non-significant group differences: $F(2,43) = .46$, $p = .63$, $\eta^2 = .02$ and $F(2,42) = 3.04$, $p = .06$, $\eta^2 = .13$, respectively. In contrast to RMSSD, however, the effects of PEP appeared to become more pronounced during the post-emotion manipulation period. This effect could also be seen in ΔPEP's significant relation to HR reactivity ($p < .001$).

***Frequency domain HRV analysis*.** Because of the necessity of extending the measuring periods for HRV frequency analyses to 120 s (Task Force, 1996), I only examined HRV effects during emotion manipulation, thus starting at the emotion manipulation period and extending 30 s into the post-emotion manipulation period. The ANCOVA for HF power data detected significant between-group differences, $F(2,45) = 4.608$, $p = .02$, $\eta^2 = .17$. Post-hoc contrasts indicated that HF power of both Guilt and Pride participants was significantly lower than that of

Neutral participants, $p = .007$, $r = .38$, and $p = .03$, $r = .32$, respectively (Figure 6A). The ANCOVA for LF power was also significant, $F(2,45) = 8.48$, $p = .001$, $\eta^2 = .27$. Post-hoc contrasts indicated that Guilt participants had reduced LF power during the emotion manipulation compared to both Pride and Neutral participants (ps = .001, rs > .47), while LF power did not differ between Pride and Neutral participants ($p = .75$) (Figure 6B). Analysis of TF power yielded between-group differences similar to LF power, $F(2,45) = 12.44$, $p < .001$, $\eta^2 = .36$, and confirmed that the Guilt group was the only experimental group with significantly reduced total heart rate variability (ps < .001, rs > .49) (Figure 6C).

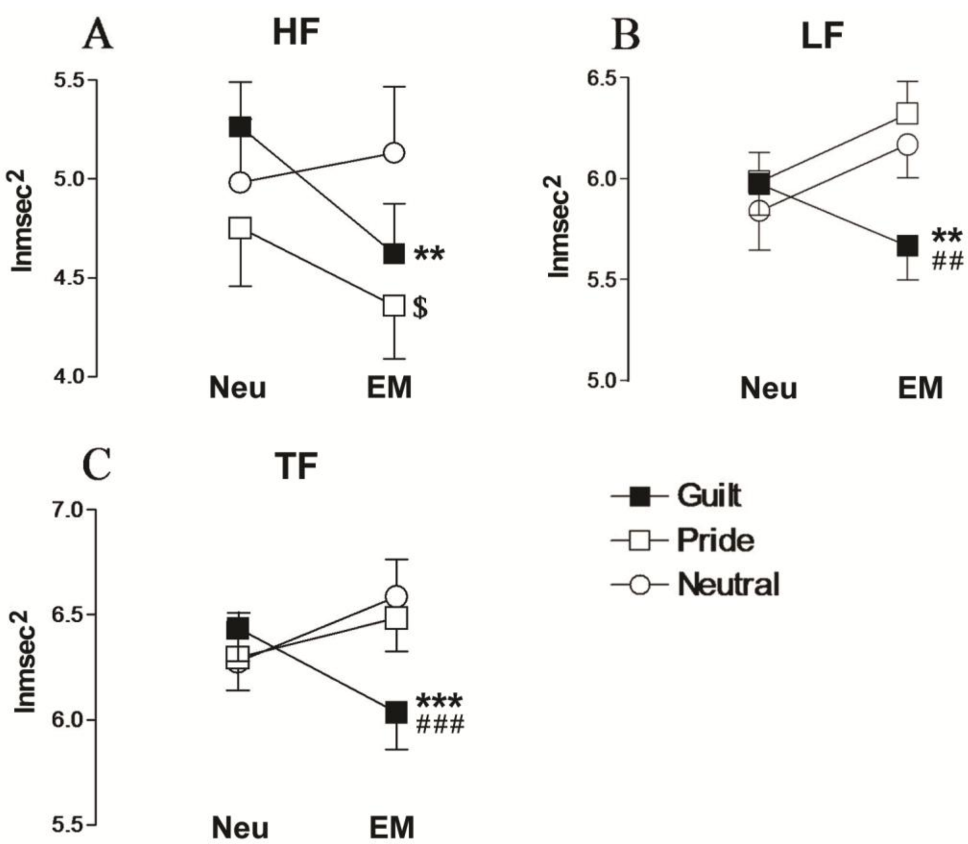


Figure 6. Power spectral densities of the three experimental conditions during the neutral and emotion manipulation periods. (A) HF, (B) LF, (C) TF.
Neu: Neutral experimental period, EM: emotion manipulation period, HF: high frequency power, LF: low frequency power, TF: total frequency power.
$^{**}p < .01$ and $^{***}p < .001$ Guilt compared to the Neutral condition. $^{\$}p < .05$ Pride compared to the Neutral condition. $^{\#\#}p < .01$ and $^{\#\#\#}p < .001$ Guilt compared to the Pride condition.

Relation between self-reported affect and physiological reactivity. To assess the relation between subjective emotion experience and physiological arousal, post-manipulation emotion ratings of guilt and pride (i.e., Emotion Time 2) were correlated with physiological reactivity scores (Table 6).

Based on the results, it could be deduced that the subjective experience of guilt was most closely related to parasympathetic withdrawal: Ratings of guilt were negatively correlated with all indices of HRV, namely, RMSSD, HF power, LF power, as well as TF power. Subjective ratings of guilt were also related to HR as well as PEP reactivity during the post-emotion manipulation period. By comparison, subjective ratings of pride were relatively unrelated to measures of physiological reactivity. In agreement with physiological data, increased pride was associated with increased RMSSD (i.e., vagal recovery) during the post-emotion manipulation period. Pride was also negatively correlated with electrodermal activity (i.e., SCL).

Table 6

Correlations Between Experimentally Induced Affect and Physiological Reactivity

Physiological Measure	Emotion Time 2: Guilt (*n* = 16)	Emotion Time 2: Pride (*n* = 16)
ΔSCL	-.12	-.42*
ΔHR EM	.26	.078
ΔHR Post EM	.57*	-.18
ΔRMSSD EM	-.36	-.07
ΔRMSSD Post EM	-.77***	.43*
ΔPEP EM	-.31	.05
ΔPEP Post EM	-.49*	.06
ΔLF EM	-.42*	.17
ΔHF EM	-.46*	-.14
ΔTF EM	-.60**	-.20

Note. Because of specific predictions regarding the direction of the correlations in the Guilt condition, 1-tailed Pearson's correlations were performed.
**p* < .05. ***p* < .01. ****p* < .001.

Individual Differences

Relation between personality measures and induced affect. Individual differences in the magnitude of experimentally-induced guilt and pride were assessed by performing partial correlations between participants' post-manipulation affect ratings (i.e., Emotion Time 2) and questionnaire scores (i.e., PANAS, BIS/BAS, and TOSCA-3), while controlling for baseline affect (i.e., Emotion Time 1). Table 7 presents the results of these partial correlations.

In the Guilt condition, participants' ratings of experimentally-induced guilt were only positively correlated with BIS, $r = .58$, $p = .02$, and negatively correlated with the BAS Drive subscale, $r = -.58$, $p = .02$. Residualized post-manipulation scores for other negative emotions (i.e., Emotion Time 1 scores covaried) indicated that BIS was not significantly related to any other negative affect: Anxiety ($r = .34$, $p = .22$), anger ($r = .07$, $p = .81$), shame ($r = .47$, $p = .08$)[5], or general negative affect ($r = .26$, $p = .34$). By comparison, BAS Drive was also negatively correlated with experimentally induced anxiety, shame and general negative affect (rs > .55, ps < .04). In the Guilt condition, BIS was therefore correlated uniquely with experimentally-induced guilt. (There was no significant correlation between BIS and baseline ratings of guilt, $r_s = .15$, $p = .30$). No other correlations between personality measures and experimentally induced guilt reached significance. In particular, the TOSCA-3 index of guilt-proneness was completely unrelated to baseline, as well as experimentally-induced, ratings of guilt (rs < .05, ps > .80).

In the Pride condition, no significant correlations were detected between experimentally-induced pride and the overall BAS scale ($r = -.34$, $p = .22$), or any of its subscales (i.e., Drive, Reward Responsiveness, and Fun Seeking). The only correlation that reached significance was between TOSCA-Beta Pride and experimentally-induced pride ($r = .52$, $p = .04$).

[5]In the case of shame, the correlation with BIS may have reached significance if the sample size was larger. This is not surprising, however, given the highly correlated nature of guilt and shame in response to self-caused wrongdoing (Schmader & Lickel, 2006).

Table 7

Partial Correlations Between Experimentally-induced Affect and Questionnaire Measures

Questionnaire Measures	Emotion Time 2: Guilt (*n* = 16)	Emotion Time 2: Pride (*n* = 16)
Positive Affect (PA)	-.12	.10
Negative Affect (NA)	.27	.48
Behavioural Inhibition System (BIS)	**.58***	.08
Behavioural Activation System (BAS)	-.33	-.34
BAS-Drive	**-.58***	-.21
BAS-Fun Seeking	.08	-.32
BAS-Reward Responsiveness	-.06	-.16
TOSCA-Shame	.25	-.07
TOSCA-Guilt	-.01	-.03
TOSCA-Alpha Pride	-.18	.42
TOSCA-Beta Pride	-.40	**.52***

Note. Correlation coefficients > .50 are in boldface. Partial correlations were performed by controlling for baseline (i.e., Emotion Time 1) levels of guilt and pride, respectively. TOSCA = Test of Self-Conscious Affect.
*p < .05.

Relation of guilt and shame proneness to BIS and BAS. Recent research by Sheikh and Janoff-Bulman (2010) identified BIS, as measured by the Carver and White BIS/BAS scales, as a unique predictor of shame proneness (measured by the TOSCA-3); they also found that BAS uniquely predicted guilt proneness (also TOSCA-3). The present analyses were conducted to determine whether these results were replicated in the current data, using questionnaire scores of all participants across experimental conditions.

In line with previous findings (e.g., Carver & White, 1994), BIS and BAS were uncorrelated in the current sample (r = -.01, p = .96). Mean responses to the TOSCA-3 measures of shame and guilt proneness, however, were highly correlated (r = .41, p = .003). The inter-correlated nature of shame and guilt proneness is also consistent with previous findings (Tangney, 1991; Tangney, Niedenthal, Covert, & Barlow, 1998; Tangney, Wagner, Fletcher et al., 1992).

To investigate the relationship between BIS and BAS and shame and guilt proneness, two linear regressions were conducted, similar to those presented by Sheikh and Janoff-Bulman (2010). Because shame and guilt proneness are highly correlated constructs, the regression analyses controlled for guilt proneness when predictors of shame proneness were assessed, and vice-versa.

For the first linear regression, the predictive value of BIS on shame proneness was tested by entering BIS, BAS, and guilt proneness simultaneously as independent variables. This analysis was statistically significant, $F(3,45) = 8.90$, $p < .001$. Consistent with Sheikh and Janoff-Bulman's (2010) findings, BIS significantly predicted shame proneness, $\beta = .33$, $p = .009$. The analysis for BAS, however, also reached significance, $\beta = -.30$, $p = .02$ (Table 8).

For the second linear regression, the predictive value of BAS on guilt proneness was tested by entering BIS, BAS, and shame proneness simultaneously as independent variables. This analysis also reached statistical significance, $F(3,45) = 4.68$, $p = .006$. Consistent with Sheikh and Janoff-Bulman's (2010) findings, BIS had no significant predictive value for guilt proneness, $\beta = .08$, $p = .6$, while the predictive value for BAS almost reached significance, $\beta = .25$, $p = .07$ (Table 8).

Table 8

Linear Regression Analyses for Variables Predicting Guilt Proneness and Shame Proneness (N = 49)

Variable	*B*	*SE B*	β
Shame Proneness			
Constant	7.00	19.05	
BIS	0.92	0.34	.33**
BAS	-.68	0.27	-.30*
Guilt Proneness	.74	0.25	.37**
Guilt Proneness			
Constant	44.01	8.13	
BIS	0.11	0.20	.08
BAS	0.28	0.15	.25†
Shame Proneness	0.22	0.08	.44**

Note. R^2 = .37 for Shame Proneness ($p < .001$); R^2 = .24 for Guilt Proneness ($p < .01$).
†$p = .07$. *$p < .05$. **$p < .01$.

Relation between BIS and BAS sensitivity and physiology. To assess the predictive power of the Carver and White (1994) BIS/BAS scales in terms of physiological parameters, simple regression analyses were performed on baseline as well as physiological reactivity (Δ) scores. Baseline physiology measures constituted the mean of the initial 5-min baseline recordings.

BIS/BAS scores were largely unrelated to baseline levels of physiological activity, that is, they were nonsignificant when predicting mean initial baseline levels of HR, RMSSD, PEP and HRV frequency measures. The only analysis that reached significance, was that of BIS and SCL ($\beta = .31$, $p = .03$). Higher BIS scores therefore predicted higher baseline electrodermal activity across participants.

BIS and BAS were also largely unrelated to physiological reactivity during the various experimental manipulations. The only significant relation existed between BIS and HR reactivity during the Guilt manipulation, such that higher BIS sensitivity predicted greater HR acceleration ($\beta = .58$, $p = .02$). None of the BAS scales were significantly related to any measure of physiological reactivity.

Discussion

Self-conscious moral emotions of guilt and pride motivate adaptive social behaviours (Leary, 2007). I created a social psychology paradigm of high ecological validity to investigate associations between guilt and pride's behavioural responses and their underlying physiological substrates. The data obtained not only confirmed effective emotion elicitation, but uncovered a *cardiac* SNS arousal pattern for guilt versus a *somatic* SNS arousal pattern for pride. This novel finding provides support for these emotions' distinct motivational functions and may be extrapolated to predict the impact of these emotions on psychophysiological processes outside the laboratory. In terms of behavioural motivation, high punishment sensitivity (as measured by Carver and White's BIS scale) was uniquely associated with greater self-reported guilt, while no relation was found between reward sensitivity (as measured by the overall BAS scale, or by any of its subscales) and self-reported pride. Finally, subjective ratings of guilt were strongly associated with parasympathetic withdrawal, as assessed by both time- and frequency-domain measures of HRV.

Did the Emotion Inductions Work?

The present investigation was designed to expand on recent trends in the field of emotion elicitation, where researchers strive increasingly to design paradigms that will elicit current emotions in psychologically real and engaging situations (Harmon-Jones et al., 2007; Herrald & Tomaka, 2002; Williams & DeSteno, 2008). Both univariate and multivariate analyses of self-report data confirmed that guilt and pride were preferentially elicited by the respective emotion manipulations: Not only did both target emotions increase significantly from baseline, but they also contributed more than any other emotion in separating the Guilt and Pride conditions from the Neutral condition. In addition, the manipulation check, where participants were forced to select the emotion they experienced most during the experiment, showed the manipulation to be very effective and emotion-specific. No explicit verbal references to guilt and pride were made during any part of the experiment, suggesting that these emotions were truly felt.

In the present investigation, I analyzed various self-reported emotions in addition to target affects through multivariate analyses. Participants were therefore required to rate several emotions that they may or may not have experienced at the time. While this multivariate approach may complicate matters, both for the participant and the experimenter, i.e., in terms of data analysis, it may also lead to invaluable insights into the subjective nature of the complex moral emotions. Emotion researchers typically try to elicit pure emotions, yet it is well established that different emotions often co-occur, and that, together, they may represent the particular emotional profiles that provide an adaptive advantage for the individual (Izard & Ackerman, 2000). For instance, in the Guilt condition of the current study, participants primarily reported feeling guilty, together with experiencing other negative emotions, i.e., anxiety, anger, shame and general negative affect (particularly disgust with situation). These emotions were also significantly correlated with experimentally-induced guilt.

For my purposes, however, it was important to distinguish guilt from the similar emotion of shame, even though these emotions have been reported to be much more highly correlated in response to self-caused than other-caused wrongdoing (Schmader & Lickel, 2006). The most salient distinction in the literature between shame and guilt, as advanced by Tangney and colleagues (2007), is the focus on self versus behaviour. Whereas a particular action is regarded as negative in guilt, the entire self is typically regarded as negative in shame (Lewis, 1971). I therefore probed for guilt-related beliefs (i.e., thoughts that one had played a causal role and should have thought, felt, or acted differently) during the post-experimental interviews. Examples of typical responses included “feeling guilty about getting the lab assistant into trouble,” “wishing I never took the money,” “feeling uncomfortable and responsible,” “wanting to write a letter to apologize,” and “feeling like I caused the situation.” It therefore appeared they were experiencing a deep regret over wrongdoing, accompanied by sincere empathy for the research assistant (who had ostensibly lost her job). Many investigators agree that notions of wrongdoing and responsibility for causing harm are critical features or determinants of guilt (Kubany & Watson, 2003; Tangney & Dearing, 2002). These results, then, are consistent with a multidimensional model of guilt (Kubany & Watson, 2003), because both negative affect and guilt-related cognitions could be identified.

Participants in the Pride condition reported feeling mostly pride and satisfaction, and these emotions were also highly correlated in this condition. Responses of pride and satisfaction are consistent with previous research, where pride has been conceptualized as the mean response to ratings of pride and satisfaction (Williams & DeSteno, 2008). Examples of typical responses after the pride induction included “feeling proud and encouraged”, “feeling that it was fun”, “feeling pleased, yet anxious at the attention”, and “feeling good and encouraged to do better.”

Although the Pride manipulation was successful, one may argue that it was less effective than the Guilt manipulation. This is evident from the discriminant function analysis, which was unable to discriminate the Pride condition from the other two experimental conditions successfully. Increases in ratings of pride, from pre- to post-emotion manipulation, were also smaller in the Pride condition than increases in ratings of guilt in the Guilt condition. Several factors may account for the fact that the pride manipulation was less emotionally intense than the guilt manipulation. First is the fact positive emotions are more difficult to elicit in the laboratory than negative emotions, a fact that relates back to the “negativity bias” of organisms (Davidson et al., 2000). This account is described in more detail in the *General Discussion*. A second possibility is that Pride participants may have been skeptical about their task performance and consequently may not have accepted the pride-inducing feedback as true. One or two participants from the Pride condition did, in fact, admit to questioning the experimental performance feedback; however, those who outright rejected the feedback as false were excluded from the sample before data analysis. A third possibility concerns the nature of the experimental manipulation itself. In contrast to the Guilt manipulation, the Pride manipulation did not involve a moral scenario affecting other people or society. It therefore did not carry the same weight or level of significance as the Guilt manipulation did. One can imagine that positive feelings of pride would be much stronger and earnest if they were in connection to a real contribution one has made to society, rather than one's performance on a task for which the outcome is probably irrelevant.

Despite being less effective than the Guilt manipulation, the Pride manipulation nevertheless provoked significant positive affect and produced a physiological pattern of arousal distinct from that of the Neutral and Guilt conditions, as elaborated in the next section. Moreover, SCL responses of Pride participants were significantly higher than those of Neutral participants, and were of a similar magnitude to those of Guilt participants. Because SCL may be interpreted as a measure of primary arousal that reflects the degree of autonomic output (Cacioppo et al., 2000), the data suggest that both emotion conditions constituted successful emotion inductions. (SCL responses in the Neutral condition may be attributed to mere interpersonal interaction with the confederates.) Because the Guilt and Pride manipulations therefore produced comparable arousal levels, the current study provided a favorable context within which to detect reliable autonomic differences between guilt and pride (Frazier, Strauss, & Steinhauer, 2004).

Somatovisceral Responses of Guilt and Pride

The physiological data supported clear between-group ANS differences: Whereas Guilt and Pride participants displayed similar levels of general arousal (i.e., SCL) compared to Neutral participants, only those in the Guilt condition displayed significantly increased HR. The fact that PEP and RMSSD were not significantly correlated in any experimental condition, however, supports the idea that cardiac responses can be mediated by simultaneous or independent changes in PNS and SNS outflows to the heart (Cacioppo et al., 1994). I endeavoured to distinguish between these effects in my physiological analyses.

The HR changes of Guilt participants were associated with reciprocal modulation of sympathetic (i.e., decreased PEP) and vagal (i.e., decreased HF power and RMSSD) outflows to the heart, with decreased PEP persisting through the post-emotion manipulation period. Analysis of self-report emotion ratings' relation to indices of physiological reactivity supported these findings: Subjective reports of guilt were most reliably associated with vagal withdrawal, as assessed by both time and frequency domain measures of HRV during EM and Post-EM, as well as PEP reactivity during Post-EM. Pride participants displayed low cardiac reactivity, with heart rate increases predominantly associated with vagal withdrawal (i.e., decreased HF power and RMSSD). Vagal unloading (i.e., RMSSD) of Pride participants, however, showed a faster recovery effect during the post-emotion manipulation period than that of Guilt participants. Subjective reports of pride were also positively associated with increased vagal power (i.e., RMSSD) during Post-EM. The modest HR increases of Pride participants are consistent with previous reports of low cardiac reactivity during pride or other positive emotions like happiness and contentment (Ekman et al., 1983; Fredrickson & Levenson, 1998; Herrald & Tomaka, 2002).

Intercorrelations between physiological reactivity measures supported the above dissociation between PNS and SNS contributions in the Guilt and Pride conditions. In the Pride condition, emotion-induced increases in HR were negatively correlated with RMSSD, reflecting the negative chronotropic effects of vagal output to the heart. In the Guilt condition, however, both RMSSD and PEP were negatively correlated with HR. Stress-induced cardiac sympathetic activation shortens PEP because of increased cardiac inotropy, giving rise to increased HR. Ancillary HR analyses, with PEP and RMSSD as additional covariates, further supported the view that reciprocal vagal withdrawal and cardiac sympathetic activation contributed to the tachycardia observed in the Guilt condition.

Interestingly, both Guilt and Pride participants displayed attenuated HF power and RMSSD during the emotion manipulation. In Pride participants, however, the decrease in HF power was accompanied by a shift to LF power (TF power remained unchanged), in the absence of significant HR increases. Consistent with previous literature, my data provided strong support for vagal modulation of LF, in that LF power was strongly correlated with HF power during all experimental conditions (Martinmäki et al., 2006). Increases in cardiac LF power are thought to reflect *somatic* SNS activation of vasomotor nerves, with the resultant vasoconstriction then leading to baroreflex-mediated vagal modulation (Brychta et al., 2007; Moak et al., 2009). Because Pride participants displayed no PEP shortening, I interpreted the reciprocal change in cardiac spectral power in these participants as a shift from respiratory-mediated to baroreflex-mediated vagal drive. This response is in stark contrast to the pronounced PEP shortening (reflecting *cardiac* SNS activation) experienced by Guilt participants. The fact that PEP and SCL were not correlated in any experimental condition further substantiated the interpretation of these variables as reflecting cardiac and somatic arousal, respectively.

To test the hypothesis that strong negative reactions take longer to down-regulate than equally strong positive reactions, I compared SCL decay (as outlined in the Method section) across experimental conditions. Because the relationship of temporal measures of electrodermal responding (e.g., recovery time) to psychophysiological processes is not as well understood as measures of amplitude or frequency of the skin conductance response, these measures are employed less frequently in research (Dawson et al., 2000). The current results, however, suggest that the measure of SCL decay holds promise for discerning sympathetic activity. In particular, my measure of skin conductance decay indicated that the average guilt response far outweighed responses in the Pride and Neutral conditions in terms of the magnitude of skin conductance level after the emotion induction. These results are consistent with previous data showing negative events to be stronger and longer-lasting than equally strong positive events (Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001). Some have argued that negative affect is up to three times stronger than positive affect, such that "one bad day must be outweighed by two or three good days in order to maintain average levels of subjective well-being" (Larsen & Prizmic, 2004, p. 44). The large variability in the SCL decay data, however, suggests that huge individual differences are at play in the down-regulation of affect.

Individual Differences

In the present investigation I examined the moderating effects of dispositional BIS and BAS sensitivities on both subjective and physiological emotion responses. In the Guilt condition, behavioural inhibition sensitivity, as measured by the Carver and White BIS/BAS scales, was uniquely associated with experimentally-induced levels of guilt, and unrelated to pre-experiment/baseline guilt. The implication of this finding is that individuals high in punishment sensitivity are more likely than those low in punishment sensitivity, to experience intense guilt in guilt-inducing situations. Experimentally-induced guilt, however, was also negatively correlated with the BAS-Drive scale, which quantifies motivation toward desirable goals. The interpretation here may be that high BAS scorers are insensitive to BIS cues of punishment and therefore less likely to experience negative affect in general. Support for this interpretation stems from the fact that scores on the BAS Drive scale were also negatively correlated with other negative emotions, i.e., anxiety, shame, and general negative affect.

The fact that BIS correlated significantly with state, but not trait, guilt is also in line with literature: The BIS/BAS scales were not originally designed to assess general affective tone, but rather to reflect behavioural tendencies in situations of threat or incentive (Carver & White, 1994). For example, given the role of the BIS in avoiding punishment via passive avoidance, someone who realizes his/her own BIS vulnerability may avoid threatening situations and in the process develop personality dimensions of constraint, e.g., Harmavoidance and Control (Fowles, 2000). In specific situations of threat, however, such a person will experience greater negative affect (in this case guilt), than someone with low BIS sensitivity. The distinction between BIS and trait negative affect was supported by the fact that negative affect (as measured by the PANAS) was not significantly correlated with scores on the BIS scale ($p = .18$).

Although guilt's role in regulating moral behaviour is widely acknowledged, contrasting views in the literature propose different functional accounts of guilt in terms of its underlying motivational orientation. Theorists from the prejudice literature argue that discrepant/prejudiced responses lead to increased guilt, which functions as a punishment cue and heightens motivation for prejudice reduction (Devine, Monteith, Zuwerink, & Elliot, 1991; Monteith, 1993). The increased self-directed negative affect (i.e., guilt) is thought to instigate a self-regulatory cycle that should cause the individual to respond more carefully in future situations (Monteith et al., 2002; Monteith, Devine, & Zuwerink, 1993). This program of research thus associates guilt with the interruption of ongoing behaviour and engagement in self-reflection. Alternative accounts of guilt, however, focus on its adaptive role as promoter of communal or prosocial behaviour (Baumeister et al., 1994, 1995; Lindsay-Hartz, 1984; Tangney & Dearing, 2002). In this theoretical view, guilt is believed to stimulate actions to make amends and restore social relationships, e.g., apologizing, confessing, and engaging in reparative behaviours, which would in turn alleviate the agony of the guilt itself (Kubany & Watson, 2003). Research focused on guilt's prosocial nature, therefore highlight its function in behavioural approach motivation.

Amodio and colleagues (2007) integrated the above accounts by separating immediate effects of guilt from its more distal implications, thus forming a dynamic model of its function. By this model, guilt first functions as a punishment cue that reduces approach motivation, after which it becomes reparative or prosocial when a suitable opportunity for amendment presents itself (Amodio et al., 2007). The current results fit well with this proposed two-stage account of guilt's motivational function. The positive correlation between BIS sensitivity and experimentally-induced guilt supports the idea that initial guilt functions as a punishment that promotes increased self-focus. The inhibitory/self-reflective nature of guilt was in fact directly evident in one case, where the participant was so lost in thought following the guilt induction that she forgot to continue with the experimental tasks until summoned to do so. Other participants also reported feeling distracted and unable to concentrate wholly on the tasks at hand after the emotion induction.

Support for the prosocial nature of guilt was also sought in the present investigation. The questionnaire that was used to assess prosocial motivation, however, did not detect significantly more prosocial responding for Guilt condition participants, than for Pride or Neutral condition participants. This result may be attributed to the fact that the questionnaire did not offer any restoration directly related to participants' current guilty feelings. The questionnaire was designed to disguise the fact that it tapped prosocial motivation (it involved voluntary participation in some unrelated HIV/AIDS studies), based on the notion that any prosocial incentives would generalize to other social domains. Findings contradicted this notion, however, which rather suggests that any prosocial behaviour that alleviates feelings of guilt should be directly related to the original transgression. This assumption is in line with evidence from the prejudice literature, where self-affirmation in areas irrelevant to the prejudice-related discrepancy, has not necessarily restored the individual's self-integrity (Dutton & Lake, 1973). Evidence for reparative behaviour, however, stemmed from the fact that 9 out of 16 Guilt participants volunteered to give back the money they had received as soon as the experiment concluded. Other prosocial motivations (e.g., wanting to write a letter of apology, and wanting to restore the research assistant's job and reputation) were also identified during the post-experimental interview, as mentioned earlier.

In terms of pride's motivation, it has recently been described as the "most important human emotion" to motivate social behaviour (Tracy & Robins, 2007a, p. 147). When we succeed in a task or meet a goal, feelings of pride are associated with increased social interaction (Noftle & Robins, as cited in Tracy & Robins, 2007), enhanced performance at subsequent tasks (Herrald & Tomaka, 2002), as well as increased efforts toward future goals despite short-term losses (Williams & DeSteno, 2008). The achievement-oriented form of pride is thus consistent with behavioural approach, or activation of movement toward positive self-rewarding goals, as measured by the BIS/BAS scales (Fowles, 1980; Gray, 1982).

In the Pride condition, however, I found no positive correlation between BAS sensitivity and experimentally-induced pride; nor was there a significant increase in HR or PEP reactivity, putative physiological markers of BAS (Arnett & Newman, 2000; Brenner et al., 2005; Fowles, 1980), compared to the Neutral condition. The most feasible explanation for this pattern of data may be that the pride manipulation did not include any specific reward incentives and that it therefore may not have produced significant changes in behavioural activation. The pride manipulation was also less intense than the guilt manipulation, and therefore may have led to only marginal increases in behavioural activation. A third possibility, however, and the explanation I favor, assumes that pride, like guilt, involves different stages. Initially, pride follows a completed success and serves to provide information about an individual's current social status (Tracy & Robins, 2007a); it is therefore unlikely to motivate immediate further action (i.e., BAS). After a while, however, a new challenge appears and it is then that 'higher' pride will activate more action than 'low' pride. This view is consistent with the fact that, unlike negative emotions, positive emotions in general are not associated with specific action programs that motivate immediate further actions (Frijda, 1986; Lazarus, 1991). I therefore propose that initial pride, as measured in the current study, does not motivate behavioural activation.

While the BIS/BAS scales pertain to *personal* reward, authentic pride is also considered an emotion that motivates prosocial behaviour, i.e., altruistic acts or reward for *society* (Tracy & Robins, 2007a). In the current study data I did find some support for the prosocial nature of pride, however, results were inconclusive possibly due to the lack of sensitivity of the prosocial measure. Nevertheless, Pride participants responded the most favourably to the prosocial questionnaire after the emotion induction. They were thus more likely than participants from the Guilt or Neutral conditions to offer their availability to participate in future research studies without further financial reward.

To assess the correspondence between behavioural ratings of approach and inhibition tendencies and physiological parameters, I conducted simple regression analyses. Consistent with previous findings (Brenner et al., 2005; Colder & O'Connor, 2004), self-reported BIS/BAS sensitivities were largely unrelated to state reactivity measures of the BIS and BAS. The only significant finding was that higher self-reported BIS, for participants in the Guilt condition, predicted greater HR reactivity during the emotion manipulation. While BIS is usually associated with electrodermal responding under conditions of frustrative nonreward (Fowles, 1988), the current finding is consistent with high autonomic arousal typically observed in high BIS scorers. For example, in the anxiety literature, children and adults with high behavioural inhibition and social anxiety tend to exhibit higher resting levels of HR and SCL, as well as greater HR and electrodermal reactivity during stress (Garrada, Connel, & Taylor, 1991; Mezzacappa et al., 1997; van Lang et al., 2007).

Consistent with previous findings, BIS and BAS sensitivities were also largely unrelated to resting levels of the physiological measures assessed (Heponiemi et al., 2004). In line with theoretical formulations of the BIS, however, self-reported BIS predicted higher resting levels of SCL. This finding is consistent with the longstanding notion that electrodermal lability is associated with trait anxiety (Fowles, Kochanska, & Murray, 2000; Pole, 2007), which falls under BIS control (Gray & McNaughton, 2000). Interpretations regarding Guilt condition participants' physiological responses and possible approach/avoidance motivations are considered in more detail in the *General Discussion*.

Measuring Guilt: Some Caveats

The finding that neither baseline nor experimentally-induced guilt was significantly related to the TOSCA-3 index of Guilt-Proneness, may *prima facie* appear contradictory. It is of paramount importance, however, that measures assessing different aspects of guilt and shame not be used interchangeably, as explained below.

Although most individuals have the capacity to experience both shame and guilt in distinct scenarios (Tangney, 1990, 1992), there are myriads of negative situations that may elicit either emotion. When faced with such an ambiguous situation, one's predisposition to respond with either guilt or shame is termed *guilt-proneness* and *shame-proneness*, respectively (Tangney, 1990). Measures that assess responses to such hypothetical scenarios, e.g., the TOSCA, thus typically assess guilt-proneness and shame-proneness. By comparison, some self-report measures have been developed to assess the frequency of actual feelings of guilt, i.e., state and trait guilt, such as the Guilt Inventory (GI; Kugler & Jones, 1992). Scenario-based and frequency-based constructs of guilt, however, are typically unrelated and may show different patterns of association with other individual difference variables (Benetti-McQuoid & Bursik, 2005). Notably, whereas TOSCA Guilt-proneness has been found to be significantly related to *prosocial*/*empathic* aspects of guilt (Tangney, 1991), the GI is more associated with *anxious* guilt (Einstein & Lanning, 1998).

Considering the above, it is readily apparent why Sheikh and Janoff-Bulman (2010) recently found a positive association between guilt and BAS sensitivity: The TOSCA, which they employed to assess guilt proneness, is associated with empathic/reparative guilt (i.e., approach motivation). Similarly, the relation they observed between BIS and shame is explained by the fact that shame-proneness, as measured by the TOSCA, has been associated with self-oriented personal distress, which is more in line with behaviour inhibition (Tangney, 1991). Despite the significantly smaller sample size of the current investigation (N = 49), compared to that of Sheikh and Janoff-Bulman (N = 120), I was able to replicate their findings regarding the predictive value of BIS on shame proneness, and BAS on guilt proneness, with reasonable accuracy.

As the present results demonstrated, however, measures of guilt and shame proneness may have little to no correspondence with the *strength* of an emotional reaction during a guilt- or shame-inducing situation. The current findings offer support for the interpretation that people more sensitive to punishment will experience greater guilt during a guilt-inducing situation.

Limitations and Conclusions

The current study's results reinforce the importance of studying emotions *in vivo* and in ecologically valid settings. A number of potential methodological limitations, however, need to be addressed. Firstly, respiration was not analyzed in the current study. I was therefore not able to distinguish between respiration-dependent and respiration-independent PNS activity (Rainville et al., 2006), or to verify that respiration fell within the HF band.[6] Houtveen and colleagues (2002), however, showed that respiratory-corrected RSA did not produce a better estimate of vagal modulation of heart rate in most stress situations. Moreover, the current experimental manipulation was not expected to alter respiratory parameters appreciably (participants neither talked nor moved), in which case uncorrected RSA may be permissible for group contrasts (Berntson et al., 1997).

Another limitation is the 500-Hz sampling rate I employed for physiological recording,

[6]It should be noted, however, that the peak frequencies of all participants, which should reflect the dominant respiratory frequencies, all fell within the .15 - .40Hz HF band.

which is below gold standard for ICG recording. Reliable impedance data, however, have been obtained even at 250 Hz using similar instrumentation. In particular, PEP estimates obtained at these lower sampling rates have been reported to have high short-term and temporal stability (Goedhart et al., 2006; Vrijkotte, van Doornen, & de Geus, 2004), as well as high heritability (Kupper, Willemsen, Boomsma, & de Geus, 2006).

A third limitation is that, in terms of the experimental design, the use of a between-subjects design over multiple experimental conditions may have introduced unwanted sources of physiological variability between participants, despite my efforts to covary out baseline differences. Within-subject designs, however, are not necessarily a good solution to this problem, because their internal validity is easily threatened (Cook & Campbell, 1979). They may, for instance, suffer from adaptation and order effects, or be entirely impractical in designs requiring deception in that the element of surprise is sacrificed (Stemmler et al., 2001).

A fourth limitation involves the fact that I investigated guilt and pride in a female-only sample. As noted earlier, I did so to avoid confounds due to possible sex-by-emotion effects. Because the issue of sex differences requires a more comprehensive discussion, however, I reserve this topic for the *General Discussion*.

A final limitation of the present paradigm was that the post-experiment prosocial questionnaire was not sufficiently sensitive to detect prosocial motivation in the respective emotion conditions. For example, in the Guilt condition, a better approach may have been to focus the questionnaire directly on behaviours that would have alleviated/rectified guilt in that particular context. Much work is still necessary to investigate prosocial responding following a guilt or pride-inducing situation, and to determine which emotion is most likely to motivate prosocial behaviour. Preliminary results from the current investigation tentatively suggest, however, that pride may be a bigger prosocial motivator than guilt.

In conclusion, the central finding of this study was the strong but dissociable sympathetic arousal during current experiences of guilt and pride. Whereas guilt was associated with reciprocal vagal withdrawal and cardiac sympathetic arousal, in pride the SNS activity was manifested by transient non-cardiac somatic arousal. Self-reported guilt was furthermore positively correlated with BIS sensitivity, supporting the conceptualization that early guilt functions as a punishment cue. In contrast, bodily arousal in early pride may be associated with a positive feeling that will induce future pride-eliciting behaviour.

STUDY 2: THE NEURAL CORRELATES OF GUILT

Over the past 20 years, numerous neuroimaging studies have been conducted on the basic emotions (e.g., happiness, sadness, fear and disgust); these have led to a better understanding of these emotions' neuroanatomical correlates and their functional organisation (see e.g., Phan et al., 2002; Phan, Wager, Taylor, & Liberzon, 2004; Vuilleumier & Pourtois, 2007). It is only in the last decade or so, however, that neuroimaging researchers have turned their attention to more complex moral cognitions and the emotions that arise in social contexts, such as guilt, embarrassment, and compassion (Immordino-Yang et al., 2009; Kédia et al., 2008; Moll et al., 2007). Given the complexity of these constructs, neuroimaging investigations are often hampered by considerable methodological challenges. Moreover, increasingly sophisticated methods of data acquisition and analysis can easily take the focus away from the elicitation side of research, making it tempting for researchers to neglect difficult issues surrounding emotion elicitation. Levenson (2003a) expressly cautions against this pitfall, arguing that emotion research data will only be as strong as the weakest link in the emotion manipulation strategy.

Moral emotions are a great deal more difficult to elicit in the confines of an MRI scanner than basic emotions are. To date, most neuroimaging studies on moral cognition have focused on the neural correlates of moral judgment or ethical decision-making (Greene et al., 2004; Greene et al., 2001; Heekeren et al., 2003; Moll, de Oliveira-Souza, Bramati et al., 2002; Sommer et al., 2010; Young & Saxe, 2008), while those that aimed to investigate specific moral emotions have primarily made use of paradigms employing emotive sentences or vignettes presented to participants in the scanner (e.g., Berthoz et al., 2002; Berthoz et al., 2006; Finger et al., 2006; Moll, de Oliveira-Souza et al., 2005; Takahashi et al., 2004, 2008; Zahn, Moll et al., 2009). Paradigms that make use of descriptive scenarios, however, focus on the perception, evaluation, and interpretation of socially relevant stimuli, rather than the actual elicitation of real emotions that motivate social behaviour (Takahashi et al., 2004). Another limitation of scenario-based paradigms is that they do not control for the influence of cultural differences between individuals, which may impact significantly on the generation of self-conscious emotions. In this regard, Haidt (2003) stressed that self-conscious emotions depend critically on aspects of social life that vary between Western and non-Western cultures. Moreover, the social background of participants, such as religion, generational cohort, and moral and political views, may all introduce confounding factors into the processing of scenarios and thereby create unwanted variability in the data (Nichols, 2002; Takahashi et al., 2008).

Perhaps the biggest confounding factor within scenario-based emotion elicitation methods, however, is that participants are not placed in emotion-evoking situations that are relevant to them personally as the causal agent within a real situation, i.e., as the person who performed the embarrassing, shameful or praise-worthy act. Rather, they are asked to *imagine* a sequence of hypothetical emotional events with themselves as the protagonist (Kédia et al., 2008). Intuitively, an imagined scenario would induce a far weaker form of emotion than a real emotion-provoking situation. From a neural perspective, Finger et al. (2006) has argued that, although one may expect significant overlap between neural activations associated with moral emotions in scenario-based methods versus real-life experiences, neural responses during the real-life experience of moral emotions would additionally include increased activity in regions associated with heightened emotional arousal, such as the amygdala, thalamus, insula, anterior cingulate cortex, and medial orbitofrontal cortex (Critchley, 2005; Critchley, Mathias et al., 2002; Patterson, Ungerleider, & Bandettini, 2002; Phelps & LeDoux, 2005; Williams et al., 2005).

In light of the above considerations, an important aim of the present investigation was to contribute to the study of moral emotions by devising an emotion elicitation paradigm of high ecological validity that would be effective inside an MRI scanner. I sought to elicit real emotions of guilt and pride that would be personally relevant to participants in order to identify neural activation associated with the experience of these emotions. To achieve this goal, I developed a neuroimaging paradigm based on a central finding in social psychology research, namely that of lingering implicit prejudice toward certain social groups in self-professed low-prejudice individuals (Devine et al., 1991; Greenwald & Banaji, 1995). A careful emotion manipulation was constructed to elicit guilt and pride in such individuals; essentially participants were led to think that they had either transgressed their personal moral values (guilt induction), or that they had responded in accordance with their personal moral values (pride induction) (Moll et al., 2008).

Implicit Prejudice as a Context for Emotion Elicitation

It is a well-established psychological fact that conscious efforts to renounce prejudice against those who are different (e.g., in their ethnicity, gender, or sexual preference) do not rid us immediately of prejudiced responses (Allport, 1954). Rather, such efforts are likely to be accompanied by a great deal of internal moral conflict between previously held stereotypic beliefs, and current efforts to eradicate all such prejudiced beliefs and reactions (Myrdal, 1944). This is particularly true in the South African context, where most people have to overcome a lifetime of prejudice-promoting socialization experiences (Williams et al., 2008).

Low-prejudice individuals are especially vulnerable to internal conflict between enduring prejudiced thoughts that are automatically activated, and consciously endorsed egalitarian values (Devine et al., 1991). Devine (1989) argued that these contradictory thoughts and feelings can coexist within the same individual and may be the reason why low-prejudice individuals often appear to have prejudiced responses that are not in line with their personal beliefs. Such discrepant responses have been demonstrated repeatedly in the literature, to the extent that people who report nonprejudiced attitudes on self-report measures frequently manifest prejudiced reactions on indirect, implicit, automatic, physiological, or neural indices of prejudice (Amodio, Harmon-Jones, & Devine, 2003; Banaji & Greenwald, 1995; Fazio, Jackson, Dunton, & Williams, 1995; Gaertner & Dovidio, 1977; Phelps et al., 2000). Adopting a personal motivation to overcome prejudiced tendencies therefore does not guarantee that someone will respond without bias across different response domains (Monteith et al., 1993). Several models have been proposed to try and explain this phenomenon (e.g., Amodio et al., 2004; Crosby, Bromley, & Saxe, 1980; Devine, 1989; Katz, Wackenhut, & Glass, 1986; Poskocil, 1977; Wilson, Lindsey, & Schooler, 2000); however, a detailed account of this literature is beyond the scope of this thesis.

Of special interest to the present investigation, however, is the fact that discrepancies between personal standards and actual responses (of prejudice) are associated with distinct affective consequences. Several theoretical frameworks have associated the transgression of personal standards with negative affect or psychological discomfort, although theories differ in their level of specificity (e.g., Duval & Silvia, 2002; Heider, 1958). Allport (1954), for example, argued that one would experience specific feelings of compunction (i.e., guilt and self-criticism) when one's *actual* reactions stand in contrast to one's personal values, and consequently, to how one *should have* behaved. Another prominent field of investigation linking negative affect with cognitive discrepancies is that of cognitive dissonance (Aronson, 1968; Bramel, 1968; Festinger, 1957). The basic tenet of dissonance theorists is that cognitive dissonance is aroused when one holds two contradictory ideas simultaneously, or when a certain behaviour threatens one's internalized standards or self-concept. In dissonance theory, however, the precise nature of the negative affect that ensues is usually expressed as a feeling of psychological discomfort or frustration, and is not usually described in terms of discrete negative emotions (Elliot & Devine, 1994; Harmon-Jones, 2000).

Higgins's self-discrepancy theory, however, provides a more articulated framework and predicts that distinct negative emotional states will arise from specific types of perceived self-discrepancies (Higgins, 1987; Higgins, Bond, Klein, & Strauman, 1986). Higgins posits that people evaluate their "*actual*" selves in terms of "*ideal*" and "*ought*" standards that are imposed by the self or some other(s), such that deviations from these create distinct negative emotions. For example, self-discrepancy theory predicts that discrepancies between actual/own versus ideal/other standards will result in vulnerability to shame and embarrassment, whereas discrepancies between actual/own versus should/own standards will result in vulnerability to feelings of guilt, self-contempt, and agitation (Higgins, 1987). Conversely, self-discrepancy theory predicts that ought discrepancies based on *others*' standards (i.e., should/other), should lead to feelings of threat and anxiety, and possibly resentment.

Self-discrepancy theory has become part of mainstream theorizing about guilt and shame and can be traced back to Freudian opposition between the id or ego, and the *ego-ideal* and *superego* (Tangney & Dearing, 2002). In neo-Freudian theory, shame ensues when one undermines the idealized standards of the ego-ideal; in contrast, guilt ensues when one violates prohibitions imposed by the superego (Piers & Singer, 1953). In a similar fashion, Higgins's (1987) theory clearly distinguishes between the type of self-discrepancy that would be associated with either guilt or shame: Shame will ensue if an act is construed as undermining that which one aspires to (i.e., ideals or goals); guilt will ensue if an act is construed as flouting a personally accepted standard for how one ought to be (i.e., norms or prohibitions) (Teroni & Deonna, 2008). Although self-discrepancy theory appears *prima facie* appealing and has been validated by some research (Monteith et al., 1993; Plant & Devine, 1998), not all experimental studies have found support for it (Ozgul, Heubeck, Ward, & Wilkinson, 2003; Phillips & Silvia, 2005; Tangney et al., 1998). Some of these claims are considered in more detail in the *Discussion* of Study 2.

Because individuals who reject prejudice have nonprejudiced values that are integrated into their self-concepts and therefore serve as a personal *should* standard (Allport, 1954; Devine, 1989; Sherman & Gorkin, 1980), it follows that low-prejudice individuals should experience specific feelings of guilt and self-criticism when they transgress their personal standards. Several theorists outside of the prejudice domain have also argued that one would feel accountable and guilty following the transgression of one's own, internalized, moral standards (Ausubel, 1955; Carver & Scheier, 1990; Hoffman, 1975). From the above, it thus appears reasonable to suggest that low-prejudice individuals view their nonprejudiced values as moral standards that they strive to uphold. Moral standards upheld in this way have been termed prescriptive moral regulation, because the focus is on what we *should* do, rather than proscriptive moral regulation, where the focus is on what we *should not* do (Janoff-Bulman, Sheikh, & Hepp, 2009). Recently, Sheikh and Janoff-Bulman (2010) found that specific types of transgressions are differentially related to judgments of guilt versus shame: Transgressions that were more likely to represent prescriptive moral violations predicted ratings of guilt, whereas transgressions that were more likely to represent proscriptive moral violations predicted ratings of shame. Taken together, research suggests that low-prejudice individuals should experience specific feelings of guilt and remorse when they transgress their own internalized moral standards.

Devine et al. (1991) tested this prediction by obtaining information from participants on how they *should* respond, according to personal standards, and how they actually *would* respond, in contact situations with Black people and homosexual men. She demonstrated that only low-prejudice participants with large discrepancies in should-would responses experienced the more specific compunction-related feelings of guilt and self-criticism. In contrast, high-prejudice participants with such discrepancies experienced only global discomfort, but not guilt. In keeping with Higgins's (1987) theory, high-prejudice participants' personal standards appeared to be based on society's conceptions to respond without prejudice, rather than on their own. Low-prejudice participants also demonstrated significantly higher levels of internalization of their personal standards than high-prejudice participants, i.e., these standards were important and central to their self-concept, and they were committed to responding in accordance with them.

In subsequent research by Monteith and colleagues (1993), they were able to corroborate these findings of differential affective reactions in response to prejudice discrepancies for high- and low-prejudice individuals. In particular, they replicated these response tendencies in situations of experimentally-induced, rather than imagined, prejudice discrepancies against gay men (Monteith, 1993). Deceiving participants into believing that they responded in a discrepant manner thus had the same effect, in terms of affective responses, as imagined scenarios.

Internal and External Motivation to Respond Without Prejudice

Motivations to respond without prejudice, however, are complex and can stem from conceptually distinct sources. People may be motivated primarily by sincere changes in their personal attitude (i.e., they may be internally motivated), or they may be motivated primarily by external pressures and social desirability concerns (i.e., they may be externally motivated). Plant and Devine (1998) devised self-report measures, the Internal and External Motivation to Respond Without Prejudice scales (IMS/EMS), to tap into these two distinct sources of motivation. They predicted that qualitatively different patterns of emotional distress would result from discrepancies of failing to live up to both one's own and others' standards, as a function of the source of motivation to respond without prejudice. Based on Higgins's (1987) self-discrepancy theory, they argued that (a) people should experience feelings of guilt and self-contempt when their actual responses violate an *own-based* should standard, but that (b) when actual responses violate a *society-based* should standard, feelings of impending punishment would ensue (i.e., fear and anxiety).

To test these hypotheses, Plant and Devine (1998) assessed should-would discrepancies as in previous studies (e.g., Devine et al., 1991), but with participants evaluating their discrepancies according to standards prescribed either by themselves (i.e., personal) or by their university campus (i.e., societal). Results for the group who assessed discrepancies from *personal* standards are particularly relevant for the present discussion, and are therefore discussed in more detail. In line with expectations, participants with high internal motivation (IMS) scores were associated with less prejudiced *should* responses. In addition, larger discrepancies between *should* and *would* responses for these individuals were associated with more negative self-directed affect (i.e., guilt). In contrast, high external motivation (EMS) scores were associated with more prejudiced would responses. Taken together, those individuals who had high IMS as well as high EMS scores were most likely to have large should-would response discrepancies, and were therefore more likely to experience elevated levels of self-directed negative affect (i.e., guilt and self-criticism). Individuals with high IMS and low EMS scores, on the other hand, had smaller should-would discrepancies and experienced only low levels of negative self-directed affect. According to Devine and Monteith (1993), the latter may be individuals who have developed the ability to consistently respond in agreement with their nonprejudiced standards (see also Plant & Devine, 2000). Conversely, individuals who are high in both internal and external motivation to respond without prejudice appear to be more likely to respond in discrepant ways from their personal standards, and therefore experience specific feelings of guilt as a result of this personal failure (Plant, Devine, & Brazy, 2003). Such individuals are concerned about others' perception of themselves, especially if their behaviour reveals prejudice.

The specific findings described above are consistent with more general theories of motivation and internalization, which state that goals and values are more successfully pursued the more internalized or self-determined they are (Deci & Ryan, 1985; Kelman, 1958; Meissner, 1981; Ryan & Deci, 2000). Self-determination theory, for example, posits that people may vary in the reasons they have for adhering to certain goals or values on a continuum of self-determination: At the lowest end, *external* reasons reflect primarily external motivations for behaviour and concerns about others' approval, while *identified* reasons constitute the opposite end and reflect highly internalized, self-determined, motivations (Deci & Ryan, 2000; Ryan & Connell, 1989). Inbetween these two extremes are moderately self-determined, or *introjected*, motivations for pursuing goals and values; these are characterized as regulation based on a combination of internal and external motivations. Introjected responses are therefore argued to be better maintained than less self-determined or external responses, yet this type of regulation is not completely integrated with the self and is therefore thought to be unstable and likely to break down in challenging situations (Deci & Ryan, 2000).

Devine and colleagues tested these predictions in the context of racial bias (Devine, Plant, Amodio, Harmon-Jones, & Vance, 2002). They obtained participants' responses on an explicit (self-report) measure of prejudice, in which responses are easy to control, as well as on two implicit measures of prejudice (an evaluative priming task and the Implicit Association Test), in which responses are more difficult to control. Devine et al. demonstrated that the most highly self-determined participants (i.e., high IMS/low EMS) were very effective at regulating their prejudiced responses on both the explicit as well as on the implicit measures. In contrast, high IMS/high EMS individuals were effective at regulating their prejudiced responses on the explicit measure, but were just as unsuccessful as low IMS (i.e., high prejudice) participants in controlling prejudiced responses on the implicit measures. Similar results were obtained by Amodio and colleagues, who demonstrated the moderating effects of internal and external motivation to respond without prejudice on physiological and neural measures assessing automatic (implicit) levels of affective race bias (Amodio, Devine, & Harmon-Jones, 2008; Amodio et al., 2003). Despite nonprejudiced self-reports, high IMS/high EMS participants exhibited larger levels of automatic affective race bias than those participants who were primarily internally motivated (i.e., high IMS/low EMS).

The cumulative program of research by Devine, Plant, and colleagues therefore suggests that, although all high-IMS individuals share internalised egalitarian beliefs, those also high in EMS are more prone to discrepancies between their actual and desired behaviours, and are therefore more likely to experience specific feelings of guilt and compunction. In addition, these individuals find it particularly hard to regulate their prejudiced responses on less controllable or implicit measures of prejudice. Although most of the research reported above pertained specifically to racial prejudice, similar findings of discrepant should-would, or explicit-implicit response biases, have also been reported in other social domains, e.g., sexuality, gender, weight, and disability (Banaji & Greenwald, 1995; Devine et al., 1991; Nosek et al., 2007; Pruett & Chan, 2006; Rojahn, Komelasky, & Man, 2008).

The Implicit Association Test

The Implicit Association Test (IAT) is a reaction-time task designed to measure the strengths of automatic associations between mental representations of objects (Greenwald, McGhee, & Schwartz, 1998). A large body of literature suggests that many cognitive processes that affect behaviour operate unconsciously or implicitly and are inaccessible to observation by the actor (Dovidio et al., 2002). Implicit cognitive measures can therefore be differentiated from self-report (or explicit) measures in that they are able to expose the strength of mental associations without depending on introspective access (Greenwald & Banaji, 1995; Greenwald et al., 2002). In this way, implicit measures can identify an individual's implicit attitude toward out-groups, regardless of whether an individual desires to expose those attitudes.

The IAT has become a popular instrument in the study of implicit social cognition due to its ease of administration, large effect sizes, and fairly good reliability compared to other implicit measures (Greenwald & Nosek, 2001). In addition, it can easily be adapted for the assessment of implicit attitudes to a wide variety of social concepts (Greenwald & Nosek, 2001; Nosek et al., 2007). The basic principle of the IAT rests on the assumption that behavioural responses (i.e., key presses) to concepts that are strongly associated in memory should be faster than to those that are only weakly associated (Greenwald et al., 1998).

While performing the IAT, one is required to rapidly categorize stimulus items (exemplars) that appear in the centre of a computer screen into one of four superordinate categories. This sorting is done through use of two response keys. The four categories usually consist of two target concepts, e.g., “insect” and “flower”, as well as two attributes, e.g., “good” and “bad”. Association strengths are assessed by comparing the speed of categorizing stimulus items in two different sorting conditions, namely congruent and incongruent. The congruent condition consists of those trials where categories that are more strongly associated for the majority of respondents share a response key, e.g., “insect” and “bad”, or “flower” and “good”. Likewise, the incongruent condition consists of those trials where categories that are less strongly associated share a response key, e.g., “insect” and “good”, or “flower” and “bad”. Respondents usually find it easier, and are therefore faster, at categorizing items where strongly associated concepts are paired (i.e., congruent trials) than when weakly associated concepts are paired (i.e., incongruent trials). The implicit response bias between the two concepts is based on this performance difference in reaction time.

A standard IAT procedure consists of five steps (or blocks) of sorting trials (see Nosek, Greenwald, & Banaji, 2005 for more details). Steps 1 and 2 involve learning how to correctly sort the concept and attribute dimensions individually. Step 3 is the first block where concepts and attributes are paired so that one key serves as the correct response for both categories (e.g., “insect or bad” and “flower or good”). In Step 4, respondents learn to switch the spatial locations of the concepts. Finally, Step 5 is the second concept-attribute pairing stage, where concepts and attributes are paired in the new configuration (e.g., “flower or bad” and “insect or good”). Steps 3 and 5 thus provide the critical reaction-time data to calculate what is typically referred to as the *IAT effect* – an estimate of the difference in association strengths (Greenwald, Nosek, & Banaji, 2003).

Because of the prominence of literature on racial prejudice, one of the most well-known IATs is one that measures bias on race (Baron & Banaji, 2006). Other popular IATs look at implicit social group attitudes, e.g., attitudes toward age, weight, sexuality and Judaism; as well as social group stereotypes, e.g., pairings between gender and career, and between race and weapons (Nosek et al., 2007).

Study Rationale

Previous research has demonstrated that low-prejudice individuals typically experience guilt when they transgress their well-internalized moral standards. Moreover, these feelings of guilt also ensue in hypothetical situations, or when participants are deceived into believing that they have transgressed their personal standards (Monteith, 1993; Monteith et al., 2002). Most relevant to the current investigation is a recent study by Amodio and colleagues (2007), who employed false physiological feedback to induce specific feelings of guilt in low-prejudice individuals. In their study, White participants were shown a series of multiracial faces (White, Asian, and Black) while their EEGs were being recorded. After the picture viewing task, participants were presented with bogus feedback about their brain-waves, which indicated negative responses toward Black faces, but not toward White or Asian faces. Both self-report data and shifts in electrical cortical asymmetry confirmed the elicitation of guilt. The context of perceived prejudice therefore holds merit as a reliable method to induce guilt.

In the present fMRI study, a modified version of the Implicit Association Test was employed as the primary instrument to elicit guilt, as well as pride, in low-prejudice individuals. Instead of calculating actual IAT results, however, the paradigm consisted of pre-programmed bogus feedback for all IATs. The IAT was considered an appropriate instrument of deception because the real test outcome is difficult to predict, and because people are not typically able to fake their responses without significantly increasing their error rates (Kim, 2001, cited in Devine et al., 2002). Pride-eliciting IATs were programmed with feedback in line with participants' self-standards, i.e., that they are non-prejudiced. Guilt-eliciting IATs were programmed with feedback contrary to participants' self-standards, i.e., that they are prejudiced. The paradigm was therefore based on the assumption that pride ensues when our actions conform to or uphold a personal moral value, whereas guilt ensues when we fail to do so or act counter to our moral values (Zahn, Moll et al., 2009). A neutral IAT condition, containing only neutral feedback, was included to serve as a control condition.

A pilot study conducted in the UCT Department of Psychology indicated that the IAT prejudice paradigm could successfully be employed to elicit both guilt and pride. Results and insights gained from the pilot study, especially in terms of selection criteria for participants, are described in more detail in Appendix D. Drawing on data from the pilot study, individuals were selected for the fMRI study based on specific demographic inclusion criteria, as well as several individual difference characteristics (e.g., low prejudice and high BIS sensitivity). These selection criteria enhanced the plausibility and success of the manipulation, because selected participants would be more likely to (a) respond in a way consistent with the pre-programmed IAT feedback, and (b) experience specific emotions of guilt, according to Higgins' self-discrepancy theory.

The current study was divided into three stages: The first two stages involved participant selection procedures, and the last stage comprised the actual fMRI study. More specifically, Stage I consisted of the initial recruitment and screening of a large pool of participants using a web-based survey, and Stage II consisted of a final screening session for selected participants from Stage I. Participants who adhered to all screening criteria from Stages I and II proceeded to Stage III.

Anticipated Results

Because previous moral emotion studies primarily made use of paradigms that assessed evaluative processes of moral emotions or moral judgments, I did not have a definitive hypothesis regarding the neural activity that would be associated with current experiences of guilt and pride. Based on theoretical assumptions, I postulated that neural activity during the guilt-elicitation would be associated with increased activity in ToM areas, including the medial prefrontal cortex (mPFC), posterior superior temporal sulcus (pSTS), temporal poles, posterior cingulate, and precuneus. Because the emotion induction was expected to be both arousing and self-relevant, I also anticipated activation in areas associated with increased emotional arousal, such as the anterior cingulate cortex (ACC), insula, and subcortical areas such as the amygdala (Phan, Taylor et al., 2004). Finally, based on data from Study 1, which suggested that the initial guilt reaction functions as a punishment cue that disrupts ongoing action, I anticipated marked ACC activation because of its strong association with BIS-related conflict-monitoring (Amodio, Master et al., 2008; Botvinick et al., 2004).

In terms of neural activation during the pride elicitation, I also expected to see significant activation in ToM areas. Additionally, I anticipated observing activation in regions implicated in positive affect, namely the mesolimbic reward pathway and its projections to the basal forebrain, e.g., the ventral striatum (Phan et al., 2002; Zahn, Moll et al., 2009). Because the pride-elicitation was expected to be less arousing and self-reflecting than the guilt-elicitation, however, I did not expect to see strong mPFC activation, nor activation in areas associated with increased emotional arousal.

Stage I: Web-based Survey

A large web-based survey was conducted to identify suitable participants for the fMRI study. Screening procedures are commonplace in emotion elicitation research, where it is of paramount importance that participants experience the specific emotions under investigation (e.g., Boiten, 1996; Reiman et al., 1997). The present measures were aimed at identifying individuals likely to have the desired affective responses following bogus feedback of implicit prejudice.

To be eligible for participation in the fMRI study, volunteers had to comply with certain demographic inclusion criteria. In addition, because the IAT emotion manipulation paradigm used in the fMRI study depended on participants holding egalitarian self-standards, participants were screened with a commonly-used measure of prejudice, namely the rating thermometer (Herek, 2000). Specifically, based on IAT topics selected for the fMRI study, participants were asked to indicate their attitude toward Black people, disabled people, homosexual people, and Jewish people. Only participants with a positive attitude toward all of these social groups were eligible. Furthermore, based on Plant and Devine's (1998) research, selected participants had to be high in both internal (IMS) and external (EMS) motivation to respond without prejudice. Finally, based on data from the IAT pilot study, participants with high BIS sensitivity were favoured because high BIS scores were positively correlated with self-reported guilt as well as pride.

The cover story conveyed to students who participated in the web-based survey was that it was a study to assess the prejudice tendencies among a university student population. Participants were also informed of a possible fMRI follow-up, and were asked to indicate their availability for this study. The only initial criterion for participation in the survey was that all participants were female. An all-female sample was chosen for the fMRI study to avoid confounds due to sex variability in emotional experience and expression, and to maximize the likelihood of intense emotional experiences (Shields, 1991).

Because the selection criteria for the fMRI study were rather stringent, a large sample of participants were recruited for the initial web-based survey. The diagram presented in Figure 7 outlines the participant flow at every experimental phase.

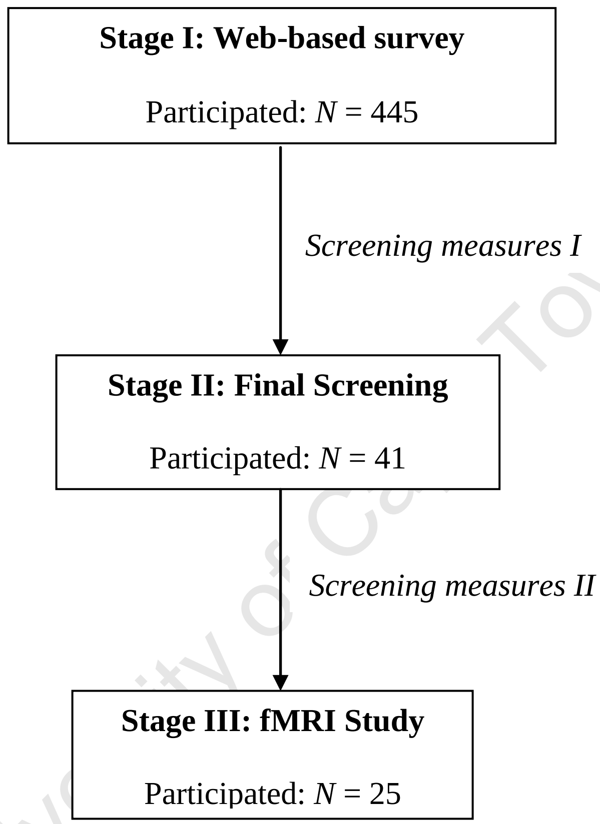


Figure 7. Flowchart detailing the number of participants at every experimental stage.

Method

Participants

A large sample of female students (N = 445), aged between 18 and 25 years, participated in the study by completing the 30-min web-based survey. Participants were recruited through web-based notification, by way of the Student Research Participation Programme's website, as well as noticeboards in the UCT Psychology Department.

All participants who completed the survey in full received course credit via the Department of Psychology's Student Research Participation Programme. All study procedures were approved by the Research Ethics Committees of the UCT Department of Psychology and the UCT Faculty of Health Sciences.

Screening Measures I

Demographic detail. Basic demographic details were requested from all participants who completed the survey (Appendix E). These included: sex, age, home language, handedness, race, religious denomination, attitude toward religion, and sexual orientation. In addition, participants were required to answer the following three questions: (i) Have you ever dated someone of the opposite race?, (ii) Do you have close gay or lesbian friends?, and (iii) Do you have close friends or family of a different race?[7]

Demographic inclusion criteria included being White, English-speaking, right-handed, non-Jewish, heterosexual, and having a neutral or positive attitude toward religion in general.

Prejudice measures. Rating thermometers (e.g., Herek, 2000) required participants to give a single attitude rating for the following social groups: Blacks, homosexuals, Jews, and disabled people. A thermometer rating of 0˚ indicated an extremely unfavorable (cold) attitude toward a particular group, whereas a rating of 100˚ indicated an extremely favourable (warm) attitude toward that group. Any rating between 0˚ and 100˚ was permissible. Thermometer ratings of ≥ 60˚ for all social groups were required to be eligible to participate in the fMRI study. An example of a rating thermometer is presented in Appendix F.

Motivation to respond without prejudice. The Internal and External Motivation to Respond Without Prejudice scales (IMS/EMS; Plant & Devine, 1998) consists of 10 items designed to assess various personal (i.e., internal) and social (i.e., external) reasons people may have for trying to respond in a non-prejudiced manner toward Blacks. Participants respond by indicating their level of agreement with each item on a scale ranging from 1 *(strongly disagree)* to 9 *(strongly agree).* The scales show good convergent and discriminant validity, and are internally consistent with alpha levels ranging from .76 to .85. Both IMS and EMS scales have also been shown to have reasonable test-retest reliability over 9 weeks, with r = .77 and r = .60 respectively (Plant & Devine, 1998).

[7]These questions were included based on results from the Pilot Study (see Appendix D).

The IMS/EMS was originally devised to assess participants' motivation to respond without prejudice toward Black people. In the present study, I adapted the original scales to assess motivations to respond without prejudice toward homosexuals, by substituting the word "Black" with "homosexual" in all statements (see Pruett & Chan, 2006, for a similar adaptation of these scales). To reduce the time taken to complete the online survey, IMS/EMS scales were not included for disabled or Jewish people.

Because those who are internally motivated to respond without prejudice should be more prone to experience specific feelings of guilt following a personal moral transgression (Higgins, 1987; Plant & Devine, 1998), it was of primary importance to select high IMS individuals. Participants with high IMS *as well as* high EMS scores were chosen, however, because such individuals have been shown to be less effective in regulating their responses on more difficult or implicit measures of prejudice (Amodio et al., 2003; Devine et al., 2002). High IMS/high EMS individuals are thus more prone to discrepant prejudice responses and to feelings of guilt.

Behavioural activation and inhibition. Individual differences in behavioural approach and inhibition sensitivity were assessed using the Behavioural Inhibition System and Behavioural Activation System scales (BIS/BAS; Carver & White, 1994). Participants with high BIS sensitivity were selected above those with low BIS sensitivity, based on findings from the pilot study.

Social desirability. The Marlowe-Crowne Social Desirability Scale (MCSDS; Crowne & Marlowe, 1960) was included as an estimate of tendency toward social desirable responding. The MCSDS is a 33-item true-false inventory that measures the degree to which individuals feel the need to describe themselves in socially desirable terms (e.g., 'I have never lied'). It therefore assesses the extent to which someone is willing to give what he/she deems to be a socially appropriate, rather than an honest, response. The set of items appearing on the full 33-item MCSDS has good internal consistency, with alpha levels in two different samples of .73 and .74, respectively (Barger, 2002). To minimize time, a 10-item short form, namely the M-C 2 (Strahan & Gerbasi, 1972), was employed in the current study. This version of the MCSDS has also been reported to have good internal consistency ($\alpha = .88$) (Fischer & Fick, 1993).

Higher scores on the M-C 2, indicating a higher anxiety about social disapproval, were preferred. Individuals high in social desirability should be more concerned with the public outcome of their test performance than those low in social desirability. They should therefore be more likely to experience positive affect in situations where they responded in a culturally appropriate manner (e.g., the pride-elicitation condition), and negative affect in situations of social disapproval (e.g., the guilt-elicitation condition).

Results

Screening Data

From the total sample of N = 445, the following groups were excluded: Blacks (n = 60); Coloureds (n = 81); Asians/Other (n = 46); homosexuals/bisexuals (n = 38); left-handers (n = 43); Jews (n = 32); atheists or participants with a negative attitude toward religion (n = 64); participants with home languages other than English (n = 6); and participants who did not want to be contacted for further participation (n = 34). After these exclusions, 154 individuals remained as potential participants for further study.

From this group, 56 with a thermometer rating below 60˚ for any of the four social groups were also excluded (Table 9). Differences in thermometer ratings between excluded and included participants were statistically significant for all social groups assessed (ps < .001).

Table 9

Thermometer Ratings for Excluded and Included Participants, Ranging from 0˚ (Very Unfavourable) to 100˚ (Very Favourable)

	Thermometer Ratings			
Selection Criteria	Black people	Disabled people	Homosexuals	Jewish people
Included (n = 98):	83.52 (13.01)	89.18 (12.68)	83.98 (11.75)	90.20 (11.60)
Excluded (n = 56):	53.84 (15.43)	71.34 (19.67)	53.21 (19.67)	63.93 (20.24)

Note. Means are presented, with standard deviations in parentheses.

Calculation of the Composite Index

For each of the remaining 98 participants, a composite index score was calculated such that higher scores indicated greater suitability for the fMRI emotion manipulation paradigm. The composite score was calculated as a % value by attributing different weights to the following questionnaire measures: IMS (*weight* = 2), BIS (*weight* = 2), EMS (*weight* = 1), and MCSDS (*weight* = 1). A Total EMS score was calculated for each participant, which consisted of the combined EMS scores for Blacks and homosexuals. Similarly, a Total IMS score was calculated for each participant. The creation of Total EMS and Total IMS scores was validated by the fact that EMS scores ($r = .75$, $p < .001$), as well as IMS scores ($r_s = .57$, $p < .001$), were highly correlated with each other.

The assigned weights were based upon theoretical inferences described earlier. In short, it was of primary importance to select high-IMS participants who were also high in EMS. Furthermore, high BIS sensitivity was imperative, because such individuals appeared to be more susceptible to the specific emotion induction based on data from the pilot study. Finally, some degree of social desirability concern was deemed favourable for the current emotion elicitation paradigm. The formula for calculating the composite index was as follows:

Composite Index (100%) = 33.3%IMS + 33.3% BIS + 16.6%EMS + 16.6% M-C 2

None of the questionnaire measures in the composite index were significantly correlated with each other (*p*s > .2), suggesting that all measures assessed largely independent constructs (Table 10). In particular, only a small negative correlation was found between the EMS and MCSDS ($r = -.12$, $p = .23$). This is in line with previous findings, and suggests that the EMS assesses a more specific concern with others' appraisal of prejudiced responses, rather than a general concern with social evaluation per se (Devine et al., 2002; Plant & Devine, 1998).

Table 10

Zero-order Correlations Among Questionnaire Measures in the Composite Index

	1	2	3
1. IMS			
2. EMS	-.01		
3. BIS	-.08	.11	
4. MCSDS	.01	-.12	.05

Note. No correlation reached statistical significance.

Stage II: Final Screening

To select participants for the fMRI study, a final screening session was arranged with suitable participants from Stage I. The main purpose of this screening was to ensure that all selected participants met safety and study criteria for the MRI scan. The session included obtaining details from participants of all current and previous medical conditions that may affect the central nervous system. In addition, a depression inventory was administered to screen out participants with noteworthy depressive symptoms. Finally, a comprehensive consent form detailing all fMRI procedures was presented to participants to ensure that they fully understood all particulars of the study before signing up for further participation.

Method

Participants

From the remaining 98 participants, those who had the highest scores on the composite index (using a cut-off of 65%), and who did not answer “yes” to *all* of the following questions: (i) Have you ever dated someone of the opposite race?, (ii) Do you have close gay or lesbian friends?, and (iii) Do you have close friends or family of a different race? were contacted for participation in this final screening session. Of the 48 participants contacted, 41 attended the screening session.

The purpose of the screening session was to exclude any participant who had a personal history of neurological or psychiatric illness and substance abuse, suffered from moderate to severe depression, used any medication that could stimulate/inhibit the central nervous system, or was unsuitable to undergo an MRI scan.

All participants received additional course credit via the Department of Psychology’s Student Research Participation Programme.

Screening Measures II

MRI suitability. A standard MRI screening (see Appendix G) assessed suitability to undergo an MRI scan. The form contained several items to ensure that participants did not have any metal objects or implants in the body, and did not suffer from any other medical conditions deemed unsuitable (e.g., uncontrolled diabetes or asthma). In addition, participants were screened for possible pregnancy, large tattoos, and claustrophobia through self-report.

Medical history and psychoactive medication. A standard self-report questionnaire (the same one as in Study 1) was used to ensure that participants did not suffer from any previously diagnosed neurological, psychiatric, cardiovascular, or substance use disorders. Participants were also required to list all current prescription medication. Those taking medications affecting the circulation, as well as any medication for depression or anxiety in the previous 6 months, were excluded. Contraceptive medication, however, was permitted.

Depression. The Beck Depression Inventory – Second Edition (BDI-II; Beck et al., 1996) was used to assess the presence of any depressive symptomatology. Because the affective characteristics associated with depression might have interfered with emotional processing, participants with depression scores ≥ 20 (the cut-off score between mild and moderate depression) were excluded. Although participants who exceeded the threshold for mild depression (i.e., a score of 14 or above) were thus admitted, it has been demonstrated that BDI scores in nonclinical populations are more likely to reflect general anxiety than clinical depression (Gotlib, 1984).

Procedure

Participants were contacted telephonically or by email to arrange an appointment for the 30-min screening session. Participation in this final screening session, as well as the fMRI study, was entirely voluntary; students were thus free to decline from further participation at any point.

Informed consent (Appendix H) was obtained according to procedures approved by the UCT Faculty of Health Sciences, and all fMRI procedures were thoroughly explained.

Results

Of the 41 participants who attended the final screening session, 12 were excluded based on the presence of metal in the body (n = 6), a high BDI-II score (n = 1), use of psychoactive medication (n = 2), a neurological history (n = 2), and being left-handed (n = 1). The remaining 29 participants signed informed consent documents and indicated availability for the fMRI study.

Screening Measure Performance

To illustrate the efficacy of the method employed for participant selection, questionnaire measure scores were compared between participants that were selected for the final screening session ($n = 41$), and those who were not selected or who did not want to participate further in the study ($n = 57$). Independent t-tests were thus performed for each screening measure.

Table 11 shows that all questionnaire means were higher for the sample that was selected for final screening compared to the one that was not. These differences reached statistical significance for BIS, EMS, and M-C 2.

Table 11

Comparison of Screening Measure Scores for Selected and Unselected Participants

	Final screening session				
Questionnaire Measure (max)	Selected (n = 41) Mean (*SD*)	Unselected (n = 57) Mean (*SD*)	t	p	Effect Size (r)
Total IMS (20)	14.67 (1.80)	14.31 (1.77)	-.99	.33	.10
BIS (28)	24.78 (2.26)	22.47 (3.39)	-4.04	<.001***	.38
Total EMS (20)	10.75 (3.20)	9.38 (3.19)	-2.09	.04*	.21
M-C 2 (10)	5.39 (1.66)	4.23 (1.54)	-3.58	.001**	.34

Note. IMS = Internal Motivation to Respond Without Prejudice; BIS = Behaviour Inhibition System; EMS = External Motivation to Respond Without Prejudice; M-C 2 = Marlowe-Crowne Social Desirability Scale (short form).
*p < .05. **p < .01. ***p < .001.

Discussion

The data presented above show that screening procedures were successfully employed to identify demographically suitable, low-prejudice participants for the fMRI study. All selected participants had a fairly positive attitude toward all social groups assessed, as measured by various rating thermometers. In addition, selected participants scored significantly higher on three other screening measures (BIS, EMS, MCSDS) compared to those who were not selected. They were thus high in behavioural inhibition sensitivity, high in social desirability concern, and high in external motivation to respond without prejudice. Participants were also high in internal motivation to respond without prejudice: The IMS mean for the current sample (*M* = 7.34) was very similar to a previous study (*M* = 7.15), where a group of high-IMS participants were specifically selected for participation (Amodio, Kubota, Harmon-Jones, & Devine, 2006).

The reason that IMS scores did not differ significantly between those selected for further participation and those not selected might be ascribed to the fact that low IMS individuals had already been excluded based on their thermometer rating scores. The IMS has consistently been shown to be highly correlated with commonly-used self-report measures of prejudice, e.g., the Attitude Toward Blacks Scale (ATB; Brigham, 1993), such that individuals who indicate significant levels of prejudice also have low internal motivation to respond without prejudice (Amodio et al., 2003; Devine et al., 2002; Plant & Devine, 1998). Consistent with these findings, Total IMS scores in the present study were significantly correlated with all rating thermometer scores (*p*s $< .001$) in the original sample of 154. Exclusion of prejudiced individuals, based on their thermometer rating scores, therefore also largely excluded those with low internal motivation to respond without prejudice.

Besides the benefit of being able to control for individual difference characteristics in the participant sample, a stringent selection process like the one described above may also be associated with some disadvantages. The most obvious disadvantage may be related to logistics, i.e., the time and effort necessary to screen such a large volume of participants. In addition, however, there may also be other drawbacks associated with screening tools. For example, by using inappropriate or inaccurate screening tools one may inadvertently select individuals who are more likely to experience a non-target emotion (e.g., shame instead of guilt). The sample may also become unrepresentative of the general population if the screening tools are too stringent, which may in turn affect the generalisability of the resulting data. Finally, self-reported data obtained by screening tools may be affected by participants' a priori theories of behaviour, rendering the data inaccurate.

In the case of fMRI investigations, however, the advantages of employing rigorous screening criteria may outweigh the disadvantages, because of the expensive nature of MRI data collection. This is especially true for emotion manipulation paradigms, where the data obtained and emotion evoked is completely dependent on the success of the paradigm for each participant.

Stage III: fMRI Study

The fMRI paradigm was based on a similar procedure as the pilot study. As in Study 1, this procedure also depended on deception to facilitate real and current emotion elicitation. During the scanning session, participants performed various different IATs, followed by pre-programmed performance feedback. The feedback they received therefore had no relation to their actual task performance; this fact was not known to participants, however. While most IAT feedback was in line with participants' beliefs about themselves (i.e., that they are non-prejudiced), some feedback suggested unconscious prejudice against a social group. Careful manipulation of IAT feedback was therefore instrumental in eliciting specific emotions of guilt and pride.

Method

Participants

Twenty-five healthy, right-handed females between the ages of 18 and 25 years (M = 19.32, SD = 1.03) participated in the fMRI study. Measures of depression and trait affective style for all participants were collected before the fMRI experiment at the final screening session [BDI- II, M = 6.64, SD = 5.02; PANAS, Positive Affect: M = 35.76, SD = 3.79; Negative Affect: M =16.96, SD = 3.98]. The duration of the entire experimental procedure, including an IAT practice session and final screening performed by the radiographer, did not exceed roughly 2hrs.

Participants received financial compensation, as well as course credit via the Department of Psychology's Student Research Participation Programme, for taking part in the study. In addition, all participants received a compact disc with a 3D image of their own brain. All study procedures were approved by the Research Ethics Committee of the UCT Faculty of Health Sciences.

Data from 3 participants were excluded before statistical analysis because of OFC signal loss (n = 1); and post-experimental interviews that suggested the emotion manipulation was not successful (n = 2), i.e., these participants did not believe that the manipulated IAT feedback pertained to their actual performance. The final sample thus consisted of 22 participants.

Experimental Design and Setting

The research design was that of a cross-sectional within-subjects social psychology experiment. This design was chosen because repeated-measures designs can control for sources of unwanted variability *between* individuals, especially in the fMRI context where the nature of the fMRI signal (no absolute zero point) may lead to confounding effects. One may thus view participants as serving as their own controls. Repeated-measures designs are routinely used in fMRI because of their greater power to detect effects, and because they are more economical in terms of cost (Britton, Phan et al., 2006; Reiman et al., 1997; Takahashi et al., 2004).

The fMRI study took take place at the Cape Universities Brain Imaging Centre (CUBIC) at Tygerberg Hospital.

fMRI Paradigm

IAT topics. The fMRI protocol consisted of six different IATs, two each of three different stimulus conditions (Neutral, Positive, and Negative). Distinctions between these stimulus conditions were created through the type of IAT topic, as well as through pre-programmed response outcomes. More specifically, the Neutral condition consisted of IATs on neutral topics (e.g., facial hair), for which no publicly endorsed right or wrong response exists; in contrast, the Positive and Negative conditions consisted of IATs on more sensitive social topics (e.g., race and sexuality). The IAT feedback, which consisted of structured sentences, was also tailored according to the specific stimulus condition: Whereas feedback in the Neutral and Positive conditions indicated no significant preference for either category, feedback in the Negative condition stated that participants' test results revealed significant prejudice against the target social group.

The two IAT topics in the Negative stimulus condition were race and disability. These IATs tested participants' unconscious prejudice against Black people and physically/intellectually disabled people, respectively. The bogus feedback participants received in the Negative condition, which suggested unconscious prejudice, was employed to induce guilt in self-proclaimed low-prejudice individuals. Race and disability, in particular, appeared to be topics well-suited for the Negative feedback condition based on findings from a mass internet study (Nosek et al., 2007). In that study, which consisted largely of a White and female sample's responses to various IATs, the race and disability IATs had among the strongest implicit-explicit effects of the social domains tested. Whereas a strong *implicit* preference for White and Abled people were thus observed from the IAT data, there was only a weak positive correlation between self-reported (i.e., *explicit*) and implicit attitudes toward these social groups (Nosek et al., 2007). Monteith and colleagues have also reported that many people can feel their racial bias when performing the IAT (Monteith, Voils, & Ashburn-Nardo, 2001).

The two IAT topics in the Positive stimulus condition were religion and sexuality. These IATs tested participants' unconscious prejudice against Judaism (versus Other Religions) and homosexual people, respectively, with pre-programmed feedback confirming participants' beliefs that they were not prejudiced against these social groups. Feedback sentences were structured to contain positive and praiseworthy personal characteristics so that participants would experience positive emotions of pride and satisfaction. Religion and sexuality IATs appeared to be well-suited for the present study's Positive feedback condition. In the Nosek et al. (2007) study, the correspondence between implicit and explicit attitudes among the social domains tested was the strongest for sexuality (r = .43), followed by Judaism (r = .38). Less discrepant prejudice responses thus existed between individuals' self-reported and implicit attitudes toward gay and Jewish people. If one assumed that participants in the present study would also have less discrepant implicit-explicit responses toward gay and Jewish people, receiving positive feedback for these IAT topics would thus enhance the plausibility of the manipulation.

Finally, the two IAT topics in the Neutral stimulus condition were facial hair and glasses. These IATs tested participants' preference for people with facial hair versus those without, and their preference for people with glasses versus those without. Although feedback in the Neutral condition also indicated no specific preference for either of these features, no positive or negative affect was expected in this condition because of the triviality of the subject matter, i.e., no specific preference was socially desirable. It was also articulated to participants before the scanning session that the outcome of these specific IATs did not really matter because there was no specific socially desirable response that applied.

IAT stimuli. fMRI stimuli for each IAT involved words and images that participants had to sort into categories as quickly as possible. Words (i.e., attributes) were categorised as either 'good' (e.g., joy, love, peace, pleasure), or as 'bad' (e.g., agony, terrible, awful, hurt); images consisted of pictures or symbols of people from each social category. For example, in the race IAT, pictures of Black and White faces were used; in the glasses IAT, pictures of people with and without glasses were used.[8]

IAT task. I programmed all IATs using *E-Prime software*, *version 1.1* (Psychology Software Tools, Inc.; Schneider, Eschman, & Zuccolotto, 2002a; 2002b). Each IAT in the current fMRI paradigm consisted of a shortened version of the standard IAT format: It included only the two critical blocks of trials (i.e., Steps 3 and 5), where concepts and attributes are paired (Nosek et al., 2005). During a pre-fMRI practice session, however, participants were given an opportunity to familiarise themselves with a more standard version of each IAT that consisted of the regular five steps. Each of the IAT blocks in the fMRI paradigm consisted of 20 trials of stimulus items that participants had to sort as quickly as possible (the design was thus self-paced). The order of presentation of congruent and incongruent blocks were furthermore counterbalanced across different IATs to reduce any order effects (Nosek et al., 2005). Correct responses had the effect of making the stimulus disappear, and were followed by a 350-ms fixation cross before the next trial. A red error sign appeared on all trials where participants failed to categorize the stimulus correctly (i.e., pressed the wrong response key), but disappeared as soon as a correct response was given.

Immediately after completing the IAT button-press task (lasting approximately 80 s), a result sentence (*Result 1*) appeared for 6 s, which indicated the outcome of the specific IAT. This was followed by one of three sentences that appeared for a further 5 s on the screen (*Result 2*), which depended on the stimulus condition: (i) “This is a regular/neutral response.”, (ii) “This is a low-prejudice/positive response!”, or (iii) “This is a high-prejudice/negative response!” Finally, three formally structured sentences appeared on a new screen for 19 s (*Result 3*). These sentences presented more extended feedback on how IAT results may be interpreted in terms of

---

[8]I acquired additional normative data for all IAT images in an independent study prior to the fMRI experiment (Appendix I). During this study, 41 participants rated all images according to three affect dimensions: valence, arousal, and dominance. The rationale for performing this study was twofold: (i) to confirm that all selected IAT stimuli were sufficiently neutral to prevent unwanted emotional reactions from being evoked during the IAT button-press task; and (ii) to exclude the possibility that picture stimuli presented across different stimulus conditions differed significantly in their affective dimensions. Data from this pilot study confirmed that selected IAT stimuli satisfied these criteria (see Appendix I for more detail); all stimuli were consequently employed in programming the six individual IATs for the fMRI paradigm.

the participant’s attitude and personality, and served to maintain/intensify the induced emotion. Figure 8 presents a diagrammatical representation of IAT events.

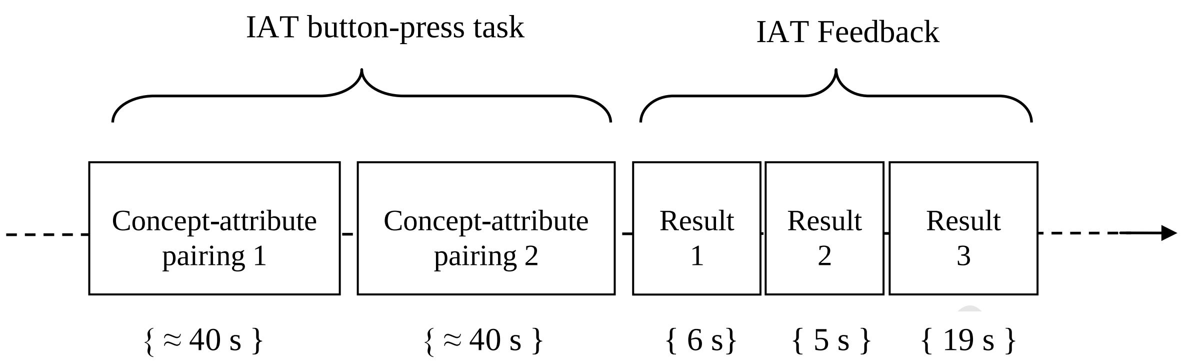


Figure 8. Illustration of the sequence of events of a single IAT during fMRI data collection.

All IAT result sentences were constructed to be similar in length, word structure, and readability across experimental stimulus conditions (see Appendix J). Results 1 and 3 were presented on a single frame in white font on a black screen, while Result 2 was either yellow, green, or red, depending on the IAT stimulus condition. Examples of IAT feedback sentences in each experimental condition are presented below:

Neutral Stimulus Condition (Glasses IAT)

Result 1: *Your data suggest little to no automatic preference for people with glasses compared to people without glasses.*

Result 2: *This is a regular/neutral response.*

Result 3: *Generally, people with no preference for people with or without glasses:*

» *Feel neutral about people who regularly wear glasses.*

» *Won’t always notice if someone changes their glasses.*

» *Often will not recall whether someone wears glasses or not.*

Positive Stimulus Condition (Sexuality IAT)

Result 1: *Your data suggest absolutely no automatic preference for Straight compared to Gay people.*

Result 2: *This is a low-prejudice/positive response!*

Result 3: *Generally, people with no automatic preference for Straight compared to Gay people:*

- » *Try to be unbiased and positive toward gay people.*
- » *Will react positively to a homosexual person in company.*
- » *Are not judgmental toward people who are gay.*

Negative Stimulus Condition (Race IAT)

Result 1: *Your data suggest a significant automatic preference for White compared to Black people.*

Result 2: *This is a high-prejudice/negative response!*

Result 3: *Generally, people with an automatic preference for White compared to Black people:*

- » *Are more afraid of Black people than White people.*
- » *Can't help having stereotypical thoughts about Black people.*
- » *Think Black people are generally less intelligent than White people.*

Subjective Responses

Emotion ratings. Subjective responses to IAT feedback were obtained during the scanning protocol to verify the elicitation of target emotional states. Participants were required to rate how much they had experienced each of 8 emotional qualities (anger, anxiety, pride, satisfaction, embarrassment, guilt, shame, and general arousal) after each IAT. Ratings were made on Visual Analog Scales (VAS), which were presented after each functional run of four different IATs. Each VAS consisted of a 10cm ruler marked from 0 (*not at all*) to 9 (*very much*), such that participants could make their ratings of experienced affects at any point along this continuous line. The presentation of all emotion items were randomized to prevent any order-effects.

Manipulation check. Upon completion of the study, participants were again shown a list with 8 possible emotions and were asked to make a forced-choice decision as to which emotion, as well as the intensity level, on a Likert-type scale ranging from 1 (*not at all*) to 5 (*very much*), they experienced most after each IAT in the scanner. For example, they were asked to indicate which emotion they experienced most prominently after the race IAT, and to rate the intensity of this emotion. The list of emotions included the same ones assessed during the scanning session; in addition, there was a neutral option if no particular emotion was felt.

Physiological Measure

During the fMRI protocol, heart rate was continuously recorded as an objective measure of emotion elicitation. Heart rate was calculated from pulse intervals recorded by a built-in peripheral pulse oximeter (Siemens Medical Systems, Erlangen, Germany), placed over participants' left ring finger. The device measures infrared light attenuation due to oxy/deoxyhaemoglobin and essentially provides a blood flow profile. Data from only 16 participants were deemed reliable due to inherent difficulties in collecting analyzable physiological recordings inside an fMRI scanner.

Procedure

On arrival at CUBIC, I informed participants of the study procedures and instructed them to read through their consent documents again. These documents assured them that they were under no obligation to participate and were free to withdraw from the study at any time. Each participant then viewed a PowerPoint slide presentation on a computer monitor, the purpose of which was to familiarize them with the Implicit Association Test. The slide presentation introduced the different IAT topics, as well as the structure of feedback that they would encounter during the scanning session. The true purpose of the experiment, however, was concealed to ensure efficient emotion elicitation and avoid demand characteristics. To heighten the realism of the emotion manipulation, I delivered the following cover story:

> *"You have specifically been selected for this study based on your explicit responses to questionnaires in the online survey. According to those questionnaires, you indicated that you have a positive attitude toward most social groups. In this study, we are interested in identifying the brain activations associated with low prejudiced behaviour. You therefore fall in our low-prejudice group and are also more likely to respond positively, or in a non-prejudiced way, to the social groups in our tasks. At*

> *a later time, we will also investigate how the high-prejudice group respond to these same tasks. Our hypothesis concerns the activity of the amygdala, a primitive organ in your brain, which we think should be less reactive in low-prejudiced than high-prejudiced individuals.*
>
> *So today we will test your implicit attitude, which is real and cannot be faked, toward these various social groups with the Implicit Association Test (IAT). The IAT can be seen as the 'gold-standard' in prejudice research, because it measures our unconscious or implicit associations, rather than our conscious or explicit thoughts about things. Although implicit responses are often similar to explicit responses, this is not necessarily always the case. However, don't worry too much about it - so far everyone has done fine on these tests."*

Participants were then given time to practise a standard version of each of the six IATs they would encounter during the scanning session. They were encouraged to respond as fast as they could without making too many mistakes. Before proceeding to the scanner, a radiographer once again ensured that all MRI safety precautions were adhered to.

Once inside the scanner, the radiographer equipped participants with earplugs (to reduce the noise of the scanner) and MRI-compatible headphones (to enable communication between the participant and the experimenter), and attached the pulse oximeter to participants' non-dominant hand. Head stabilization was achieved using foam padding within the head cradle, and participants were instructed to avoid head movements during the scanning procedures. They were then instructed to relax and close their eyes for a 9-min period during which localization and structural MRI data were obtained. Immediately thereafter, participants proceeded with the IAT protocol.

All stimuli were viewed through a mirror system mounted to the head coil, with a data projector connected to a desktop outside the scanner sending stimuli to this mirror system. *E-Prime* software was used to display stimuli during the scanning session and record participants' reaction times and subjective response data via a response box. Two variants of each of the six different IATs were presented, resulting in 12 IATs in total. The experimental design was blocked, so there were thus four IATs/blocks for each of the three stimulus conditions (i.e., Neutral, Positive, and Negative). All IATs were pseudo-randomly arranged to form three functional runs per scan, such that each run contained at least one Neutral, one Positive, and one Negative stimulus condition (see Figure 9).

To prevent any order effects, three counter-balanced presentation orders were devised so that every 4th participant received the same order of IATs (Figure 9). The only other constraint in these IAT sequences was that a Neutral condition IAT was always presented first.

IAT Code:	Facial hair	Glasses	Sexuality	Religion	Race	Disability
	1	**2**	**3**	**4**	**5**	**6**

	Run 1	Run 2	Run 3
IAT Sequence 1:	**1 4 2 5**	**6 3 5 1**	**3 2 6 4**
IAT Sequence 2:	**2 3 6 4**	**1 5 2 4**	**5 1 3 6**
IAT Sequence 3:	**2 5 3 6**	**4 1 3 5**	**2 6 4 1**

Figure 9. Diagrammatical representation of the various IAT sequences that were employed in the fMRI protocol. Neutral stimulus condition: facial hair and glasses IATs; Positive Stimulus condition: sexuality and religion IATs; Negative Stimulus condition: race and disability IATs.

Each IAT consisted of a self-paced button-press task followed by result sentences (i.e., Feedback) that were of fixed duration (illustrated in Figure 10). All IATs were interleaved with 20-s fixation periods, while an initial 20-s fixation epoch served as the baseline. During presentation of IAT feedback, participants were instructed to simply pay attention, and were told that no response was required. Upon completion of a functional run, participants rated how they felt while viewing their IAT results during each of the preceding IATs. They performed emotion ratings for each presented emotion on the visual analog scales. This was followed by a 20-s rest period before the next functional run commenced. The maximum time spent in the scanner never exceeded approximately 50min.

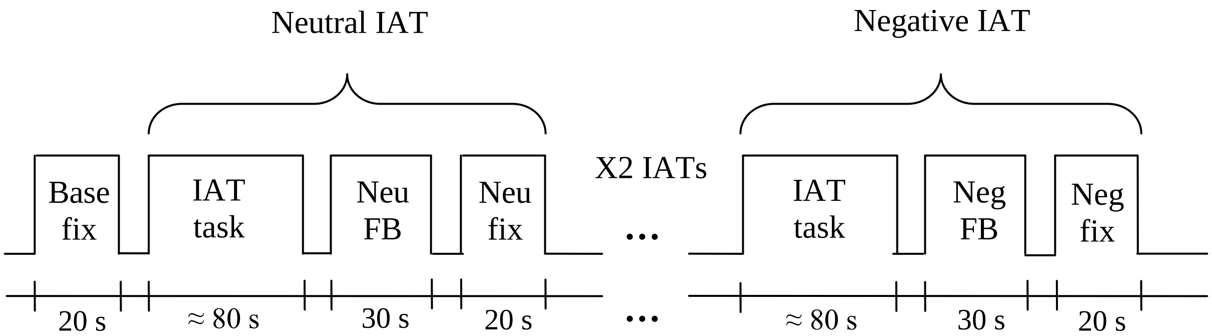


Figure 10. Block design paradigm as an example of one functional run in the fMRI study. Base: baseline, fix: fixation, Neu: Neutral, Neg: Negative, FB: feedback

Upon completion of the scanning procedures, participants were asked to describe how they felt about their IAT results, and if any results contradicted their own expectations. They were then asked to indicate which particular emotion they experienced most during each specific IAT feedback condition (this served as the *Manipulation Check*). Post-experimental emotion ratings, however, were obtained *before* the real purpose of the investigation was overtly stated. Through careful enquiry, I also determined whether the participant was suspicious of any part of the experimental procedure, or believed that IAT results were pre-programmed. Each participant was then thoroughly debriefed as to the actual intention of the investigation and great care was taken to ensure that all participants left in a positive frame of mind. In particular, participants were informed that all IAT results – those indicating prejudice as well as those indicating no prejudice, were pre-programmed, and that deception was necessary to ensure that the emotions they experienced were elicited naturally.

fMRI Image Acquisition

A 3.0 T Siemens Allegra head-only system (Siemens Medical Systems, Erlangen, Germany) was used for data acquisition. At the outset of each session, a high-resolution T1-weighted anatomical scan (3D mprage sequence, 1x1x1 mm^3 resolution, 160 slices ~ 9min) was obtained. Whole-brain functional echoplanar image volumes (T2*-weighted epi-BOLD, TE = 30ms, TR = 2s, 34 interleaved slices, 3 mm thick, gap 0.9 mm, 200 x 200 mm field of view, matrix 64 x 64, orientation approximately parallel to the AC-PC line) were then acquired using blood oxygen level dependent (BOLD) contrast while participants performed the task. Functional runs started with four steady state volumes that were discarded prior to data analysis to allow for T1 equilibration effects. A series of approximately 300 volumes of measurement were acquired during each of the three functional runs (~10 min each). Each run comprised 4 different IAT topics, with at least one IAT from each of the three stimulus conditions (i.e., Neutral, Positive and Negative), plus one extra. This design, instead of one featuring only 3 IATs per run, i.e., 1 x Neutral, 1 x Positive and 1 x Negative, was preferred to make the order appear less apparent to participants. The IAT task was set up to trigger automatically at the start of each measurement. At the termination of each run, however, the scanner was stopped manually because the design was self-paced.

Statistical Analysis

Behavioural and subjective emotion data. To investigate possible differences in task performance during fMRI data acquisition, mean response times to IAT stimuli were compared between the three different stimulus conditions.

In terms of subjective response data, a one-way repeated measures ANOVA was performed on arousal ratings to detect significant differences in subjective arousal between the baseline and IAT stimulus conditions. More specific emotion ratings acquired during the scanning procedure were examined using a 3 (stimulus condition: Neutral, Positive, Negative) x 4 (emotion type: social positive affect, social negative affect, anger and anxiety) repeated-measures ANOVA. Emotion ratings for IATs from the same stimulus condition were thus grouped together. The social positive index consisted of mean ratings for pride and satisfaction, while the social negative index consisted of mean ratings for guilt, embarrassment, and shame. These emotion indices were created to reduce the number of dependent variables for data analysis. Post-hoc contrasts were employed to detect significant changes in subjective responses between stimulus conditions.

One-way (emotion type: social positive affect, social negative affect, anger, and anxiety) repeated-measures ANOVA and post-hoc analyses tested whether target emotions were elicited most strongly in each stimulus condition. Paired *t*-tests were used to detect significant increases/decreases in ratings of target emotions from baseline. Finally, manipulation check data, from the post-experimental interviews, were analyzed to verify that target emotions were elicited.

Physiological data. Heartbeats were counted during the first 20 s of each IAT feedback condition and compared to the baseline fixation period (20 s) of the same run. HR change scores were thus computed by subtracting the mean baseline fixation value from the mean value for each specific feedback condition. These physiological change scores were then averaged within *a priori* stimulus conditions for each participant, first within each run, and then across all runs. Changes in HR for each feedback condition were examined using a one-way repeated-measures ANOVA and post-hoc analyses.

fMRI data analysis. I performed all fMRI data analyses with Brain Voyager QX, version 2.1.2 (Brain Innovation B.V., Maastricht, The Netherlands). Functional images were corrected for different slice acquisition times using cubic spline interpolation, based on information about the TR (2000ms) and the order of slice scanning (ascending, interleaved). Images were also temporally smoothed with a high-pass filter (GLM with Fourier basis set) of two cycles. 3D motion correction was applied to detect and correct any inter-scan head movement by spatial alignment of all volumes of a participant, relative to the first volume of the functional run. Translation and rotation parameter estimates were inspected and all data with motion exceeding 3 mm displacement or 3.0° rotation were rejected. Two runs, one from participants 15 and 22, each exceeded these motion criteria and were excluded from subsequent analyses. Each participant's functional data sets were co-registered with her high-resolution anatomical MRI, rotated into the AC-PC plane and spatially normalized to the standard Talaraich brain template to allow for group analysis. Normalised 4-D volume time course (VTC) files were thus created for each functional run of participants.

Functional data were analyzed using standard neuroimaging methods on the basis of the General Linear Model (GLM) (Friston et al., 1994). Seven predictors corresponding to known experimental blocks were entered into the model: IAT button-press task, IAT feedback predictors (i.e., Neu FB, Pos FB, and Neg FB) and fixation predictors (i.e., Neu Fix, Pos Fix, and Neg Fix). In addition, the six motion parameters, calculated during realignment, were also used in the model as additional predictors of no interest to remove variations in signal due to movement artifacts. In order to account for haemodynamic delay and dispersion, each of the predictors were derived by convolution of an appropriate boxcar waveform with a double-gamma haemodynamic response function. The main blocks of interest were those corresponding to IAT feedback during which emotion elicitation was expected. Feedback periods lasted 30 s and consisted of Results 1, 2, and 3 (see Figure 8).

To assess relative differences in activation between stimulus conditions, a random effects analysis was performed. This analysis provides a good generalization to the population from which data are obtained, because the error variance is estimated for each condition of interest across subjects, rather than across scans (Holmes & Friston, 1998). At the first level, parameter estimates (beta coefficients) of block-related activity were obtained for each voxel in the brain. Statistical maps of the *t*-statistic were then created for each contrast of interest and analysed at the second level using a random-effects ANOVA. To assess specific condition effects, the contrasts of Neg FB minus Neu FB (i.e., guilt) and Pos FB minus Neu FB (i.e., pride) were performed. In addition, because the effects of an emotion manipulation may continue beyond stimulus presentation (compare Study 1, i.e., the post-emotion manipulation period), contrasts between condition specific fixation periods (i.e., Neutral, Positive, and Negative) were also performed.

For the whole-brain analysis, significant clusters of activation were detected using a statistical threshold corrected for multiple comparisons using the False Discovery Rate method $p(\text{FDR}) < .10$ (Genovese, Lazar, & Nichols, 2002), and cluster size threshold of 2 contiguous voxels. To confirm whether the percent signal increase in activated areas extended over the entire measuring period, event-related averaging plots were inspected.

To confine the cortex volume to relevant brain areas and increase the power of the analysis, a mask was created from the contrast Neg FB minus IAT (button-press task) in the random effects analysis, using a corrected threshold of $p(\text{FDR}) < .10$, and cluster size threshold of 8 contiguous voxels. (Similar activations were obtained for the Pos FB minus IAT contrast. Because these areas of activation were slightly smaller, however, the Neg FB minus IAT contrast was used to create a mask of maximum size.) Areas of decreased activation (e.g., motor areas activated by the IAT button-press task) were deleted from the mask. The mask was then applied to the cortex volume and the random-effects analysis was again performed. For the mask analysis, significant clusters of activation were detected using a corrected threshold of $p(\text{FDR}) < .10$, and cluster size threshold of 2 contiguous voxels. Areas of significant activation were labelled using a human brain atlas (Duvernoy, 1999).

I also conducted correlational analyses to demonstrate a more direct link between regional brain activations, as uncovered by the whole-brain as well as the mask analyses, and subjective emotional reports. Following suggestions by Vul et al., (2009), subjective emotion ratings were correlated with fMRI brain activations of regions that were independently selected, based on functional constraints. That is, analyses were performed to detect linear relationships between mean β-values (i.e., effect sizes representing the percent signal changes) of those regions of increased activation uncovered during the whole-brain and mask analyses, and subjective emotion ratings for each subject.

Results

Behavioural Data: Response Time

Behavioural data from IATs were analyzed to verify that participants were appropriately engaged in the task. During the scan, participants responded via button press to sort different IAT images and attributes into appropriate categories. Table 12 presents the mean response times to IAT stimuli across stimulus conditions.

All response time data complied with assumptions of parametric data. The one-way repeated measures ANOVA for response time detected significant differences between stimulus conditions, $F(2,42) = 5.36$, $p = .008$, $\eta^2 = .20$. Post-hoc contrasts indicated that response times to stimuli from the Negative IAT condition were significantly faster than response times to stimuli from the Neutral ($p = .009$, $r = .53$), as well as the Positive ($p = .04$, $r = .44$) IAT conditions. Response times did not differ between the Positive and Neutral IAT conditions ($p = .19$, $r = .23$).

Table 12

Mean Response Times to Stimuli in Different IAT Stimulus Conditions

	IAT button-press sorting task		
	Neutral Mean (*SD*)	Positive Mean (*SD*)	Negative Mean (*SD*)
Response time (ms)	724.97 (135.76)	711.64 (119.96)	691.93 (108.66)

Subjective Emotion Data

Within-scan emotion ratings. One participant did not complete within-scan emotion ratings correctly, and so her data were excluded from all subsequent analyses. Subjective emotion ratings obtained during the IAT protocol are presented in Table 13.

The one-way ANOVA for subjective arousal detected significant differences between baseline and experimental conditions, $F(3,60) = 25.88$, $p < .001$, $\eta^2 = .56$. Post-hoc contrasts indicated that arousal ratings were significantly lower than baseline in the Neutral and Positive IAT conditions ($p < .001$, $r = .70$ and $p < .01$, $r = .54$), while they were significantly higher than baseline in the Negative IAT condition ($p = .002$, $r = .63$). Arousal ratings for the Negative IAT condition were also significantly higher compared to that of the Neutral and Positive IAT conditions (*p*s < .001, *r*s > .75).

Table 13

Emotion Ratings From 1 (Not At All) to 9 (Very Much) at Baseline, and After the Three Experimental Stimulus Conditions (N = 21)

	Baseline Mean (*SD*)	Neutral IAT Mean (*SD*)	Positive IAT Mean (*SD*)	Negative IAT Mean (*SD*)
Arousal	4.63 (1.55)	3.39 (1.46)	3.76 (1.75)	5.96 (1.31)
Pride	4.49 (1.50)	4.90 (1.52)	6.04 (1.39)	2.34 (0.86)
Satisfaction	4.68 (1.34)	5.15 (1.25)	6.06 (1.45)	2.38 (0.98)
Anxiety	5.17 (1.50)	3.33 (1.66)	3.34 (1.46)	5.83 (1.66)
Anger	1.83 (1.43)	2.17 (1.45)	2.10 (1.29)	5.19 (1.60)
Embarrassment	2.49 (1.72)	2.34 (1.32)	2.11 (1.28)	6.60 (1.29)
Guilt	2.14 (1.61)	2.15 (1.34)	2.20 (1.40)	6.73 (1.76)
Shame	2.06 (1.62)	2.22 (1.30)	2.26 (1.30)	6.43 (1.47)

Note. Emotion ratings were obtained during the scan, after each completed run of IATs.

The two-way (stimulus condition x emotion type) repeated-measures ANOVA detected a significant main effect for stimulus condition, $F(2,40) = 58.79$, $p < .001$, $\eta^2 = .75$. Post-hoc contrasts revealed significant differences in emotion ratings between the Negative and Neutral IAT conditions, $F(1,20) = 80.38$, $p < .001$, $r = .89$, as well as between the Negative and Positive IAT conditions, $F(1,20) = 75.04$, $p < .001$, $r = .88$, but not between the Positive and Neutral IAT conditions (p = .15, r = .31).

The main effect for emotion type was also significant, $F(1.75, 35.08) = 10.95$, $p < .001$, $\eta^2 = .35$, indicating that quantitative emotion ratings differed significantly across the four emotion types. Because Mauchly's test was significant and the assumption of sphericity was therefore violated, $\chi^2(5) = 30.11$, $p < .001$, the degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = .59$).

Finally, there was a significant interaction between stimulus condition and emotion type, $F(2.34, 44.77) = 79.13$, $p < .001$, $\eta^2 = .80$. (Because Mauchly's test was again significant, $\chi^2(20) = 93.47$, $p < .001$, a Greenhouse-Geisser correction was applied, $\varepsilon = .37$). Emotion ratings therefore varied depending on the stimulus condition (Figure 11).

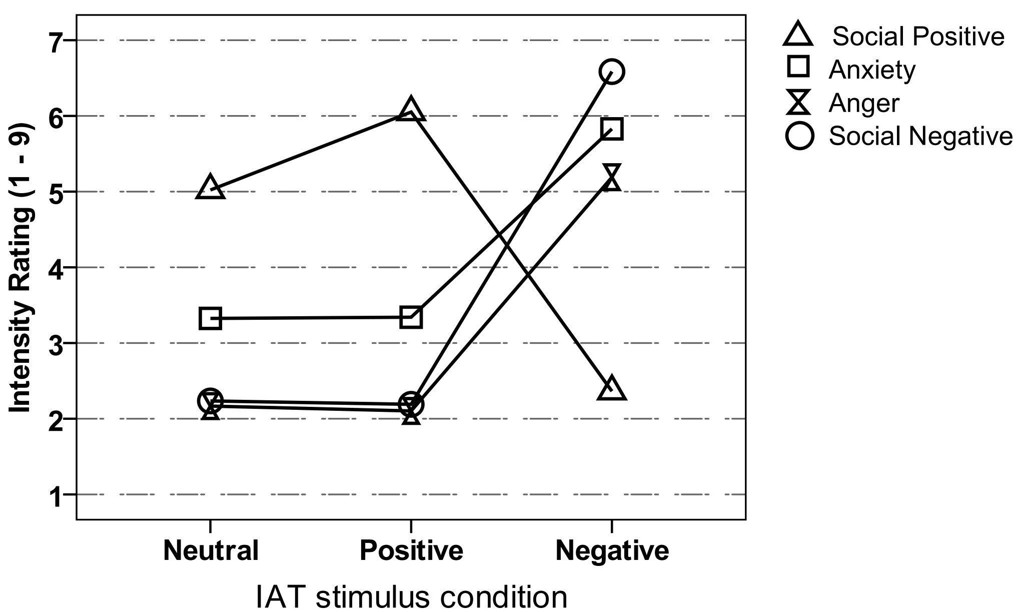


Figure 11. Interaction graph for the two-way repeated-measures ANOVA.

To break down the interaction, several post-hoc comparisons were performed. Significant interactions were revealed when comparing the Negative IAT condition to the Neutral IAT condition for social negative affect (i.e., guilt, embarrassment, and shame) compared to anxiety ($p < .001$, $r = .76$), anger ($p = .001$, $r = .67$), and social positive affect ($p < .001$, $r = .93$). Looking at the interaction graph (Figure 11), these effects reflect the fact that social negative affect increased significantly more than anxiety, anger, and social positive affect in the Negative

IAT condition compared to the Neutral IAT condition. Very similar interaction effects were found in the Negative IAT condition compared to the Positive IAT condition, i.e., social negative affect again increased significantly more than anxiety, anger, and social positive affect (ps $\leq$ .001, rs > .66).

Remaining contrasts revealed significant interactions when comparing the Positive IAT condition to the Neutral IAT condition for social positive affect (i.e., pride and satisfaction), compared to social negative affect (p = .001, r = .64), anxiety (p = .001, r = .64), and anger (p = .001, r = .66). In terms of the interaction graph (Figure 11), these effects reflected that social positive affect increased significantly more than social negative affect, anxiety, and anger in the Positive IAT condition compared to the Neutral IAT condition.

One-way repeated-measures ANOVAs confirmed that target affects were elicited more strongly than any other emotion in each stimulus condition. For the Negative IAT condition, the one-way repeated-measures ANOVA for emotion type was significant, $F(2.04, 40.87) = 47.12$, $p < .001$, $\eta^2 = .70$ (Greenhouse-Geisser correction, $\varepsilon = .68$). Bonferroni post-hoc tests were selected because they are likely to control the Type I error rate (Maxwell, 1980). These indicated that ratings of social negative affect were significantly higher than anger (p = .001, r = .63) as well as social positive affect (p < .001, r = .90), while the comparison between social negative affect and anxiety almost reached significance (p = .054, r = .54) (Figure 12).

Similarly, for the Positive IAT condition, the one-way repeated-measures ANOVA for emotion type was significant, $F(1.61, 32.21) = 56.27$, $p < .001$, $\eta^2 = .74$ (Greenhouse-Geisser correction, $\varepsilon = .54$). Bonferroni post-hoc tests indicated that ratings of social positive affect were significantly higher than anxiety, anger, and social negative affect (ps < .001, rs > .80). The one-way repeated-measures ANOVA for the Neutral IAT condition was also significant, $F(1.78, 35.62) = 32.08$, $p < .001$, $\eta^2 = .62$ (Greenhouse-Geisser correction, $\varepsilon = .59$). This condition had a similar emotion profile to that of the Positive IAT condition, because social positive affect in this condition was also rated as significantly higher than all other emotions (ps < .01, rs > .65) (Figure 12).

From the above analyses, it could be concluded that social negative affect was rated the most highly in the Negative IAT condition, while social positive affect was rated the most highly in both the Positive and Neutral IAT conditions. To explore these results further, paired t-tests were performed for each of the individual affect ratings in the social positive and social negative affect indices, from baseline to the respective IAT condition.

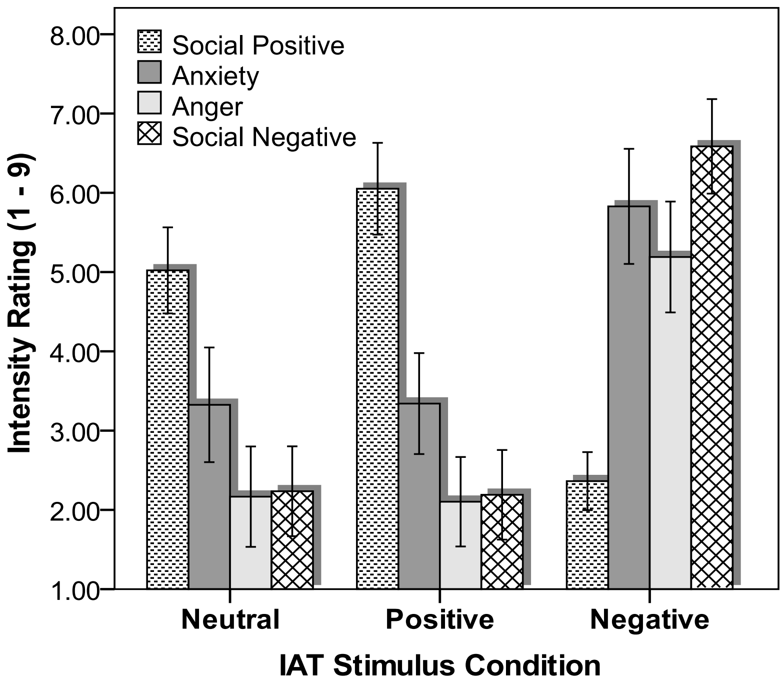


Figure 12. Means for ratings of affect in each IAT stimulus condition during the fMRI scan. Error bars represent the standard error of the mean.

In the Negative IAT condition, guilt (M = 6.73, SD = 1.76), embarrassment (M = 6.60, SD = 1.29), and shame (M = 6.43, SD = 1.47) all increased significantly from baseline, $ts(20) >$ 8.7, $ps < .001$, $rs > .85$. T-test results could therefore not differentiate between these highly correlated negative emotions ($rs > .65$, $ps \leq .001$). In the Negative IAT condition, pride was negatively correlated with guilt (r = -.66, p = .001), embarrassment (r = -.72, $p < .001$), and shame (r = -.50, p = .02). Similarly, in the Positive IAT condition, pride (M = 6.04, SD = 1.39) and satisfaction (M = 6.06, SD = 1.45) increased significantly from baseline, $rs(20) > 3.8$, ps = .001, $rs > .65$, and were also significantly correlated with each other (r = .75, $p < .001$). In the Neutral IAT condition, however, there were no significant increases in either pride or satisfaction from baseline ($ps > .16$, $rs < .30$).

These results indicated that participants reported significantly increased guilt, embarrassment, and shame in the Negative IAT condition, while they reported significantly increased pride and satisfaction in the Positive IAT condition. Although ratings of pride and satisfaction were also higher than other emotions in the Neutral IAT condition, there was no *increase* in positive affect from baseline in this condition.

Manipulation check. The manipulation check after the scanning procedure was employed to verify that target emotions were successfully elicited during the fMRI paradigm. Participants were required to state which emotion (from a list of 8), and with what intensity (from 1 to 5), they experienced most during the feedback period for each IAT. The manipulation check was especially important because within-scan emotion ratings could not distinguish between guilt, shame, and embarrassment in the Negative IAT condition, or between pride and satisfaction in the Positive IAT condition. Participants were required to state which emotion(s) they experienced predominantly during IAT feedback. Differences between shame, guilt, and embarrassment were explained to participants where necessary.

Table 14 gives the percentages of participants who reported experiencing a particular emotion during the presentation of each IAT's feedback. Emotion ratings for IAT topics in the same stimulus condition were very similar and therefore averaged together. When considering the percentage values, it is clear that participants felt mostly neutral in the Neutral IAT condition, but experienced predominantly guilt in the Negative IAT condition, and predominantly pride and satisfaction in the Positive IAT condition. Mean intensity ratings (from 1 to 5) for these emotions were as follows: Guilt (M = 3.88, SD = 0.85), pride (M = 3.83, SD = 0.66), and satisfaction (M = 3.57, SD = 0.64). Participants furthermore reported that they found the IAT feedback convincing and were generally unsuspicious in terms of their IAT results (only 2 participants indicated some suspicion during the experimental paradigm).

Taken together, subjective emotion data confirmed that participants experienced mostly social negative affect and anxiety during the Negative IAT condition, and social positive affect during the Positive IAT condition. Post-experimental manipulation data verified that, when asked to select one emotion, most participants indicated that guilt was the emotion they felt most in the Negative IAT condition. For the Positive IAT condition, participants selected either pride or satisfaction as the emotion they experienced most. During the Neutral IAT condition, participants mostly reported feeling neutral, although some also felt satisfied.

Table 14

fMRI Manipulation Check Data: Experienced Affect in Response to Various IAT Stimulus Conditions (N = 22)

Stimulus Condition	IAT Topics	Emotion: Neutral	Anxiety	Satisfied	Pride	Guilt	Embarrassed	Shame	Anger
Neutral	Facial Hair	68%	4.5%	27%	-	-	-	-	-
	Glasses	54%	-	45%	-	-	-	-	-
	M	**61%**	2.2%	36%					
Positive	Sexuality	-	-	54%	72%	-	-	-	-
	Religion	-	-	50%	63%	-	-	-	-
	M			**52%**	**67.5%**				
Negative	Race	-	4.5%	-	-	100%	9%	-	13.6%
	Disability	-	9.0%	-	-	91%	27%	13.6%	9.0%
	M		6.75%			**95.5%**	18%	6.8%	11.3%

Note. Predominant emotions in each IAT stimulus condition are in boldface. *M* = mean.

Physiological Data

Figure 13 presents changes in HR across all three experimental runs in response to the different IAT feedback conditions. A linear valence effect could be identified, such that the highest HR reactivity was associated with the Positive IAT feedback condition, while the lowest HR reactivity was associated with the Negative IAT feedback condition. This pattern of HR responses was also observed in each individual run of the fMRI paradigm (Table 15).

The one-way repeated measures ANOVA for HR reactivity detected statistically significant differences, $F(2,30) = 5.74$, $p = .008$, $\eta^2 = .28$. Post hoc contrasts indicated that HR reactivity during the Negative IAT feedback condition was significantly lower than HR reactivity during the Positive ($p = .003$, $r = .68$), as well as Neutral ($p = .036$, $r = .51$) IAT feedback conditions. HR reactivity during the Positive and Neutral IAT feedback conditions did not differ significantly ($p = .48$, $r = .18$).

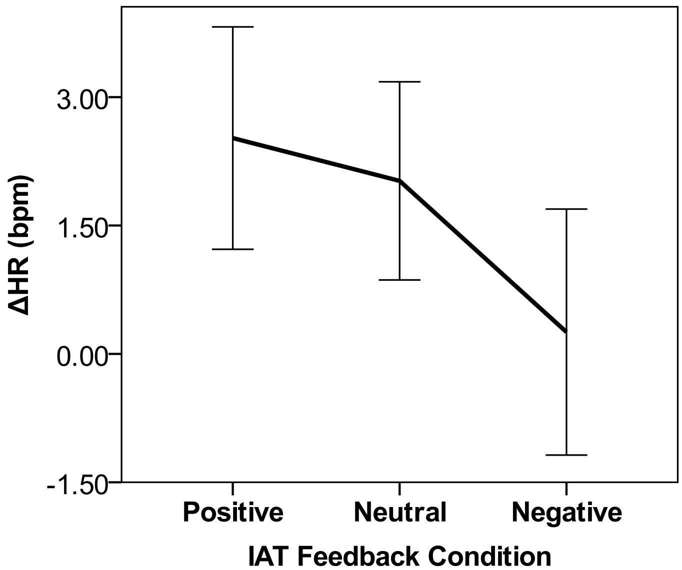


Figure 13. Heart rate change scores (from baseline fixation) for the different IAT feedback conditions.

Table 15

Means and Standard Deviations of Participants' Heart Rate Responses to Positive, Neutral, and Negative IAT Feedback Conditions (N = 16)

	IAT Feedback Condition		
ΔHR (bpm)	Positive Mean (*SD*)	Neutral Mean (*SD*)	Negative Mean (*SD*)
Run 1	2.50 (4.22)	2.50 (5.29)	0.53 (4.64)
Run 2	2.91 (5.57)	1.78 (4.00)	1.36 (4.40)
Run 3	2.16 (3.51)	1.78 (3.84)	-1.13 (3.35)

Zero-order correlations were performed to assess the association between HR reactivity and subjective emotion reports. After the removal of two outliers, there was a significant positive linear relation between ratings of guilt and overall HR reactivity in the Negative IAT condition (r = .56, p = .03). Ratings of pride were not significantly associated with HR reactivity in the Positive IAT condition (r = .30, p = .28).

fMRI Activations

Whole-brain analysis. The random-effects ANOVA yielded statistically significant results of which the specific contrast effects are described below. At a corrected threshold of p(FDR) < .10, the contrast for guilt (i.e., Neg FB – Neu FB) produced significant activation in two regions within the superior frontal gyrus/dorsomedial PFC, as well as in supragenual ACC (supraACC), and pregenual ACC (pACC) (Table 16 and Figure 14). Event-related averaging plots for these areas confirmed that the signal increase extended over the entire 30-s feedback period, thus starting at Result 1 and continuing to the end of Result 3 (Figure 15). No significant activations were observed for the Neg Fix – Neu Fix contrast.

At a corrected threshold of p(FDR) < .10, the contrast for pride (i.e., Pos FB – Neu FB) did not reveal any significant activation clusters. Even at a more relaxed threshold of p < .001 (uncorrected), no significant activation clusters could be distinguished for either the Pos FB – Neu FB contrast, or the Pos Fix – Neu Fix contrast. These results implied that the Positive and Neutral IAT conditions were too similar to yield significant differences in neural activation. Because the contrast for pride did not result in significant brain activation, a conjunction analysis, to assess common areas activated by the guilt and pride conditions, could not be conducted.

Table 16

Regional Activations for the Guilt Condition (Neg FB > Neu FB)

Brain regions	L/R	Coordinates			Voxels	*Max t*
		x	*y*	*z*		
1. Superior frontal gyrus and supraACC	L	-4	28	30	408	6.04
2. Superior frontal gyrus	L	-4	43	18	153	5.48
3. pACC	L	-7	34	15	184	4.96

Note. Talaraich coordinates and t-score refer to the peak of each brain region. A corrected threshold of p(FDR) < .10 and cluster size threshold of 2 contiguous voxels was employed. SupraACC = supragenual ACC; pACC = pregenual ACC.

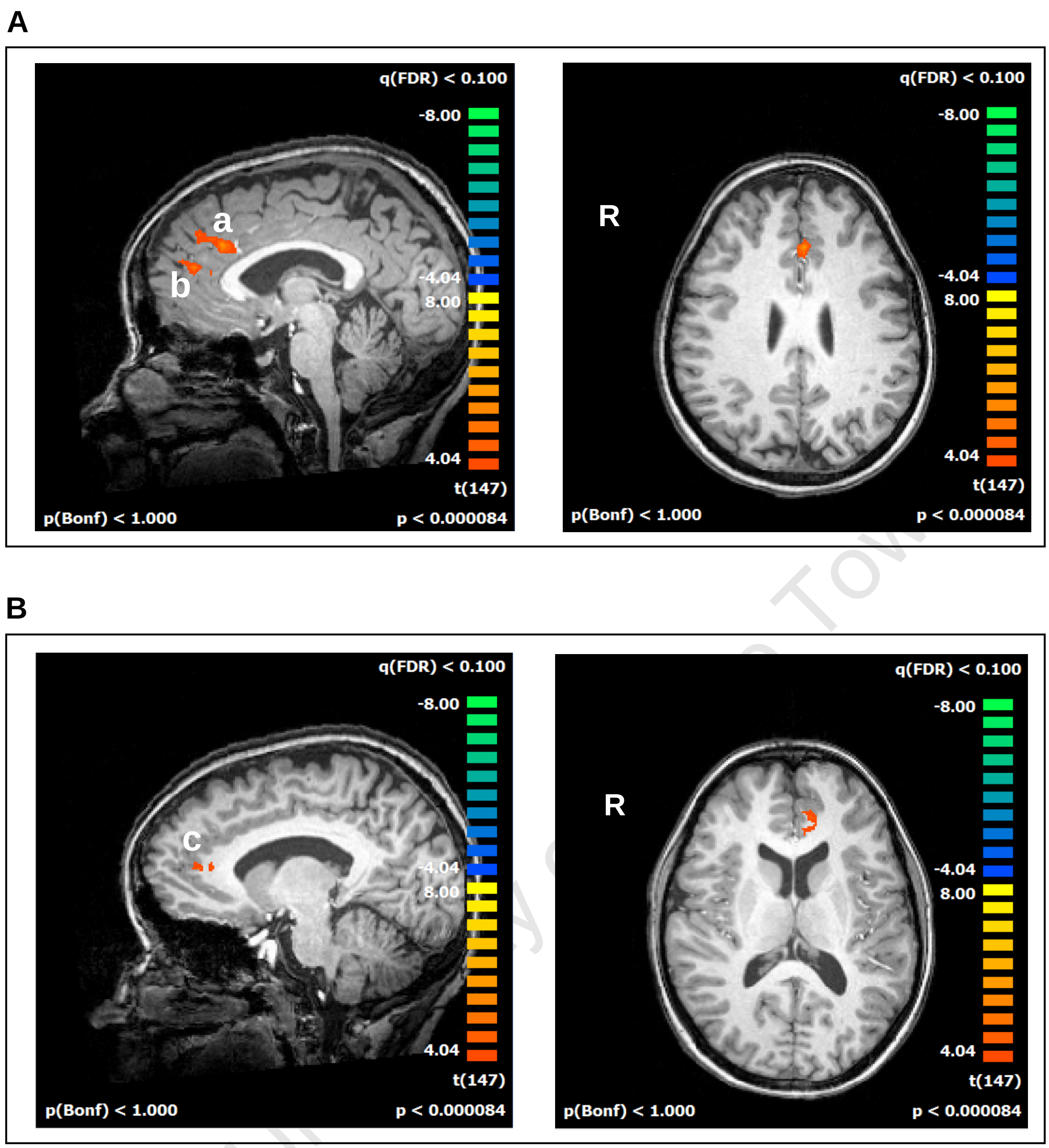


Figure 14. Images showing brain activations for guilt in the Neg FB – Neu FB contrast. Activated areas were in the A) superior frontal gyrus/supragenual ACC (label a) and superior frontal gyrus (label b), and B) pregenual ACC (label c). Significant differences were recognised at a corrected threshold ($t > 4.05$; $p < .0001$) and cluster size threshold of 2 contiguous voxels.

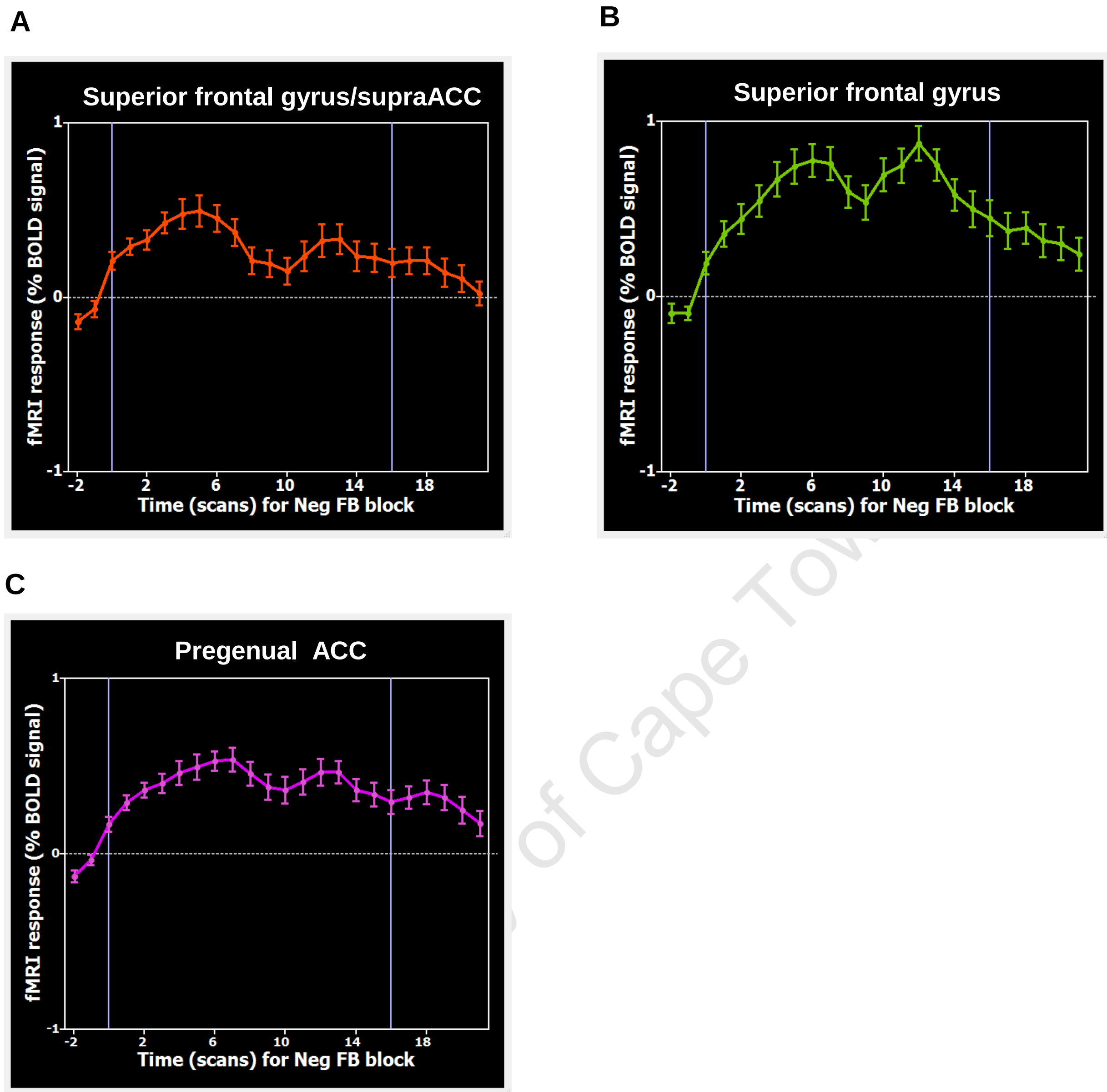


Figure 15. Event-related averaging plots indicating the percent signal change for the regions activated by the Neg FB – Neu FB contrast. A) superior frontal gyrus and supragenual ACC (-4,28,30), B) superior frontal gyrus (-4,43,18), and C) pregenual ACC (-7,34,15) (see Table 16).

Mask analysis. The mask that was created from the contrast Neg FB – IAT included most of the PFC, large areas of the temporal lobes bilaterally, the posterior cingulate and precuneus, as well as subcortical areas (Figure 16). Areas that were more active during the performance of the IAT (i.e., those showing deactivation), were not included in the mask. The final mask consisted of ~194 924 voxels.

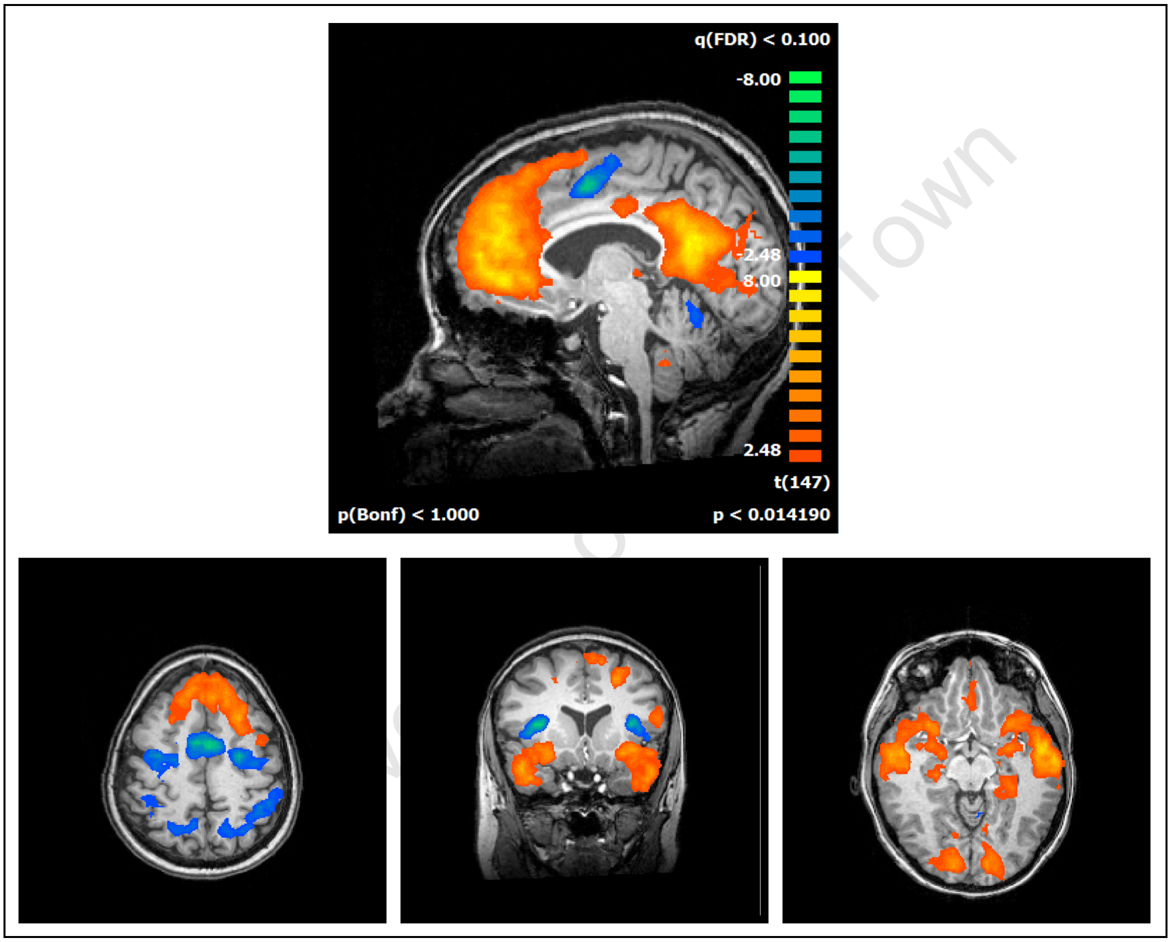


Figure 16. Brain activation clusters for the contrast Neg FB – IAT, which were used to create a mask. Significant differences occurred at a threshold of p(FDR) < .10, and cluster size threshold of 8 contiguous voxels. Areas in blue showed deactivation, and were excluded from the mask.

The random-effects analysis in the reduced cortex volume produced highly significant results. The contrast for guilt (i.e., Neg FB – Neu FB) now yielded significant activation in several areas, including the mPFC and ACC, left anterior insula and lateral orbital gyrus/ventrolateral PFC, right posterior insula, right hippocampus, right mediodorsal thalamus, precuneus, and posterior cingulate gyrus (Table 17 and Figure 17).

The contrast for Neg Fix – Neu Fix also produced significant activation in several areas, including the mPFC, left and right anterior middle temporal gyri, left posterior middle temporal gyrus (lateral and mesial), left and right angular gyri (also known as the posterior STS), posterior cingulate, left OFC, and left DLPFC (Table 17).

The Positive IAT condition compared to the Neutral IAT condition still did not result in any significant activation clusters.

Table 17

Mask Analysis: Regional Activations for the Negative IAT Condition (i.e., Guilt) Relative to the Neutral IAT Condition

Brain regions	L/R	Coordinates			Voxels	*Max t*
		x	*y*	*z*		
Neg FB minus Neu FB						
mPFC and ACC	L/R	-4	28	30	9450	6.04
Posterior insula	R	32	-20	6	137	4.20
Hippocampus	R	23	-23	-12	136	4.54
Thalamus	R	3	-23	3	55	4.82
Anterior insula and ventrolateral PFC	L	-36	19	0	683	4.43
Posterior cingulate	L/R	-4	-20	33	1665	4.30
Precuneus	L/R	-1	-68	21	824	3.74
Precuneus/posterior cingulate	L	-13	-47	30	73	3.57
Neg Fix minus Neu Fix						
mPFC (mostly left)	L/R	-7	25	51	5362	5.02
Middle temporal gyrus (anterior)	R	53	7	-21	469	4.06
Middle temporal gyrus (anterior)	L	-52	4	-21	313	4.94
Middle temporal gyrus (posterior, mesial)	L	-49	-35	-3	453	4.65
Middle temporal gyrus (posterior, lateral)	L	-61	-41	-9	420	3.89
Angular gyrus (posterior STS/TPJ)	L	-40	-64	30	1906	5.29
Angular gyrus (posterior STS/TPJ)	R	44	-59	30	330	3.74
Posterior cingulate	L/R	-4	-17	33	310	3.52
Middle frontal gyrus (DLPFC)	L	-47	1	39	195	3.66
Ventrolateral PFC	L	-49	25	-3	397	3.72

Note. Talaraich coordinates and *t*-score refer to the peak of each brain region. Significant activations are reported at a corrected threshold of $p(\text{FDR}) < .10$, and cluster size threshold of 2 voxels (Neg FB – Neu FB), and 6 voxels (Neg Fix – Neu Fix), respectively.

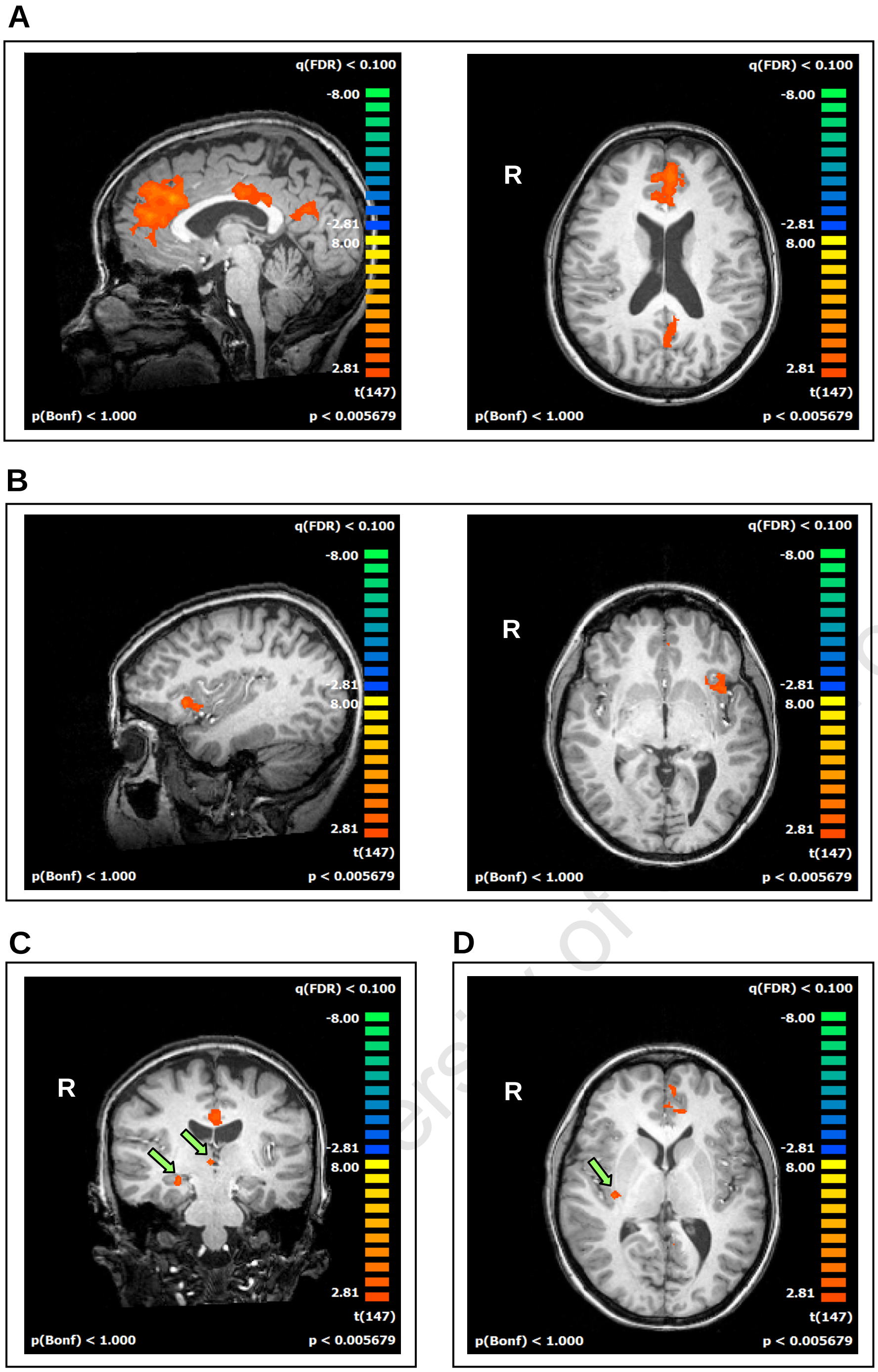


Figure 17. Mask analysis: Images showing brain activation in the Neg FB – Neu FB contrast. Activated areas included the A) mPFC, posterior cingulate and precuneus, B) left anterior insula and ventrolateral, PFC C) right hippocampus and thalamus, and D) right posterior insula. Significant differences were recognised at a corrected threshold of *p*(FDR) = .10 and cluster size threshold of 2 contiguous voxels.

Correlational analysis. Pearson's correlation coefficients were computed to assess linear relationships between subjective emotion ratings and fMRI signal changes in regions of significant activation during the Negative IAT condition (i.e., guilt). All parameter estimates for Neg FB – Neu FB, as well as subjective ratings of guilt, were normally distributed (assessed using Kolmogorov-Smirnov tests). Regions of interest uncovered during the whole-brain and mask analyses were also correlated with each other to obtain a better understanding of the linear relationships between brain areas during a current guilt experience. Because so many correlations were computed, only those significant at the 1% level (i.e., $p < .01$) were interpreted.

Highly significant negative correlations were observed between subjective ratings of guilt and the degree of activation in the pACC ($r = -.70$, $p < .001$), as well as the right posterior insula ($r = -.57$, $p = .007$) (Figure 18). Other negative emotions (anxiety, embarrassment, and shame) were also negatively correlated with activity in the pACC (*r*s < -.63, *p*s < .01); however, guilt was the only negative emotion that correlated significantly with activity in the right posterior insula (at the 1% level).

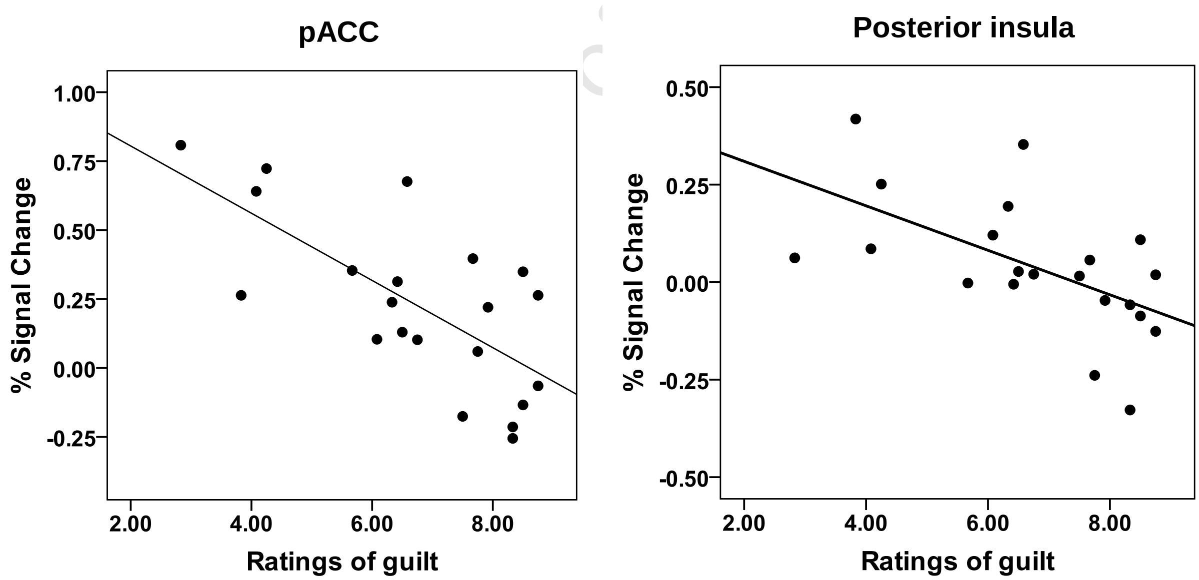


Figure 18. Plots and regression lines of correlations between emotion ratings and fMRI signal changes in specific brain regions. Self-ratings of guilt for the Negative IAT condition were negatively correlated with the degree of activation in the pACC (x = -7, y = 34, z = 15; 184 voxels, $r = -.70$, $p < .001$), and the right posterior insula (x = 32, y = -20, z = 6, 137 voxels, $r = -.57$, $p = .007$).

As can be seen in Table 18, fMRI activations indicated that the pACC and right posterior insula were also significantly correlated with each other ($r = .58$, $p = .006$). Interestingly, ratings of pride, which decreased during the Negative IAT condition, were linearly related to activity in the right posterior insula ($r = .64$, $p = .002$).

Other brain regions that showed significant linear relationships were the mPFC, which was positively correlated with the precuneus and posterior cingulate regions (rs > .55, ps < .01), while the precuneus and posterior cingulate were significantly correlated with each other ($r = .75$, $p < .001$) (Table 18).

Table 18

Zero-order Correlations Among Regions of Significant Activation for the Negative IAT Condition (i.e., Guilt) Relative to the Neutral IAT Condition

	1	2	3	4	5	6	7	8	9	10
Whole-brain analysis										
1. Superior frontal gyrus/supraACC										
2. Superior frontal gyrus	.37									
3. pACC	.16	.22								
Mask analysis										
4. mPFC and ACC	.71***	.70***	.58**							
5. Posterior insula (R)	-.08	.21	.58**	.18						
6. Hippocampus (R)	-.36	-.14	.34	-.02	.21					
7. Thalamus (R)	.04	.08	.30	.27	.50	.26				
8. Anterior insula and ventrolateral PFC (L)	.31	.38	.27	.53	.19	-.34	.14			
9. Posterior cingulate	.20	.46	.38	.57**	.16	.38	.15	.25		
10. Precuneus	-.02	.44	.26	.40	.08	.40	.10	.22	.75***	
11. Precuneus/posterior cingulate	.11	.58**	.27	.35	.24	-.21	.12	.35	.29	.47

Note. supraACC = supragenual ACC; pACC = pregenual ACC; R = right hemisphere; L = left hemisphere.
**p < .01. ***p < .001.

Individual differences: BIS/BAS. As in Study 1, subjective emotion ratings of guilt were correlated with participants' scores on the BIS/BAS scales to determine the relations between BIS sensitivity, BAS sensitivity, and experimentally-induced affect. In addition, participants' BIS and BAS scores were correlated with the degree of activation in the pACC, and in the right posterior insula.

Participants' scores of BIS and BAS were uncorrelated in the fMRI participant sample ($r = -.09$, $p = .69$). Consistent with findings from Study 1, higher BIS scores were positively associated with subjective ratings of guilt, although this correlation did not reach significance. Significant negative correlations, however, were detected between BIS scores and areas that were associated with subjective ratings of guilt, i.e., the pACC and right posterior insula (Table 19). BIS was not significantly associated with activity in any other brain region.

In contrast, BAS was significantly correlated with self-reported guilt, such that lower ratings of BAS were associated with higher ratings of guilt ($r = -.62$, $p = .003$) (Table 19). In addition, BAS was positively correlated with signal changes in both the pACC ($r = .68$, $p = .001$), and the right posterior insula ($r = .48$, $p = .03$), but not with activity in any other brain region. Taken together, these findings suggest that, irrespective of BIS sensitivity, those participants with lower BAS sensitivity were more likely to experience intense guilt.

Table 19

Correlations Between BIS and BAS Sensitivity and Self-reports of Guilt, as Well as Signal Changes in Specific Brain Regions

	Guilt	pACC	Posterior insula (R)
BIS	.24	-.44*	-.44*
BAS	-.62**	.68**	.48*

*$p < .05$. **$p < .01$.

Discussion

I conducted an fMRI study to examine the neural correlates of feelings of guilt. In a novel departure from previously published studies, the paradigm employed elicited this moral emotion in real-time and as a salient affective state relevant to the participant. To this end, the paradigm made use of preprogrammed feedback of either high or non-existent prejudice on an implicit attitude task designed to elicit guilt (as well as pride) in low-prejudice individuals.

As hypothesized, the guilt condition produced increased BOLD responses in areas implicated in the neural substrates of mentalising, as well as in anterior paralimbic structures associated with increased emotional arousal. In particular, anterior medial frontal activation (BA 9/32) during guilt suggests that participants engaged in heightened self-reflection, whereas dorsomedial (BA 8/9) and ventrolateral PFC (BA 47) activation may be associated with the processing of socio-emotional cues that signal punishment or unacceptable social behaviour. Moreover, prominent conflict-related supragenual ACC activity suggested that acute guilt is associated with the interruption of ongoing behaviour. Contrary to predictions, no significant temporal lobe activations were detected during the guilt condition; these were observed during the Negative fixation period directly following the emotion induction, however. A noteworthy finding was the significant negative association between self-reports of guilt and the degree of activation in the pregenual ACC (pACC), which may point to individual differences in the regulation of negative affect. Finally, there were no significant fMRI activations associated with the pride condition, suggesting that this emotion manipulation was either not intense enough, or not sufficiently different from the Neutral IAT condition, to evoke specific neural activations.

In summary, the direct, real-time elicitation of guilt was associated with a network of neural activation in areas consistently implicated in previous imaging studies of moral emotion elicitation. These areas have, however, never before all been activated by a single elicitation paradigm in a single study. Moreover, the distributed pattern of neural areas activated during guilt gives credence to the notion that guilt is a multi-faceted construct, including self-reflection, heightened arousal, mentalising, the interruption of ongoing behaviour, and affect regulation, which together may serve to guide and direct moral behaviour in complex ways.

Efficacy of the Emotion Elicitation Paradigm

In order to make any inferences about brain activations detected in the current investigation, it was necessary to confirm that target emotions of guilt and pride were elicited successfully. Unlike most previous fMRI designs, where participants were only asked to rate their emotional experiences post-hoc (e.g., Berthoz et al., 2002; Kédia et al., 2008; Moll et al., 2007; Takahashi et al., 2008; Takahashi et al., 2004; Zahn, Moll et al., 2009), emotion ratings in the current study were obtained *during* the scanning procedure. Subjects were furthermore required to rate several emotions in addition to target affects, and were questioned about their subjective experiences, in order to obtain rich qualitative data on the nature of the elicited affect.

Subjective emotion reports indicated that the paradigm was successful in eliciting specific emotions of guilt, and to a lesser extent pride: Not only did target emotions increase significantly from baseline to the respective IAT stimulus conditions, but they were also the most salient emotions experienced during the Positive and Negative IAT conditions, respectively. In the Negative IAT condition, high ratings of guilt were also associated with high ratings of embarrassment and shame. Embarrassment and shame were thus also felt, despite the fact that participants reported almost uniformly (during the manipulation check) that guilt was their overriding feeling during the Negative IAT feedback period. Similar findings, namely of elevated ratings of guilt, shame, and embarrassment in response to stories portraying victim-based moral transgressions, were reported by Finger et al. (2006). Finger et al., however, indicated that the presence of an audience resulted in significantly higher ratings of embarrassment and shame in response to moral transgressions, while it did not impact significantly on ratings of guilt. Ratings of embarrassment and shame in the present investigation may therefore have been augmented by a ‘public’ factor, because participants were aware that their performances were being monitored outside the scanner. The co-occurrence of guilt with other negative emotions is considered in more detail in the *General Discussion.*

In the Positive IAT condition, participants reported mostly increased pride and satisfaction, similar to emotional reports from Study 1, and post-experimental reports verified that participants experienced heightened pride in response to Positive IAT feedback. In the Neutral IAT condition, however, participants also rated pride and satisfaction as the emotions they experienced most, even though pride and satisfaction did not increase significantly from baseline in this condition. While most participants reported feeling neutral in response to Neutral IAT feedback, quite a few also reported that they did not distinguish considerably between the Positive and Neutral IAT conditions, and felt pleased with any feedback that was non-prejudiced, because that is how they believed they “should be.” The Neutral IAT condition was thus perceived by some participants as more positive than intended. The similarity in subjective emotion profiles between the Positive and Neutral IAT conditions probably prevented a clear distinction between these two conditions; the paradigm was therefore unsuccessful in identifying neural activations associated with pride.

Because an important aim of the current investigation was the creation of an improved, ecologically valid, method for the elicitation of guilt and pride within an fMRI environment, the employed paradigm warrants closer inspection. In terms of pride, several authors have argued that we may feel pride when we uphold or act in accordance with our personal moral values (Haidt, 2001; Hume, 1739/1984; Moll, Zahn et al., 2005; Zahn, Moll et al., 2009). This response is also predicted by Duval’s objective self-awareness theory, which states that congruity between self and some personal standard should result in positive affect (Duval & Silvia, 2002; Duval & Wicklund, 1972). In terms of theoretical considerations, it is therefore unclear why the paradigm did not succeed in producing strong positive affect. The most feasible explanation may be that participants experienced the Positive IAT condition as a cognitively pleasing condition, rather than a true positive emotion-evoking condition (Levenson, 2003).

In terms of guilt, because experimental reports for self-discrepancy theory have been inconsistent, it was necessary to investigate some of those claims in more detail. To recap, Higgins’s self-discrepancy theory predicts that discrepancies between actual/own versus ought/own standards will result in vulnerability to feelings of guilt, whereas discrepancies between actual/own versus ideal/other standards will result in vulnerability to shame (Higgins, 1987, 1999). Because participants in the current study were selected to have high internal motivation to respond without prejudice, their integrated, nonprejudiced values are believed to have served as a personal *should* standard (Devine et al., 1991), making them more prone to experience guilt, instead of shame, when transgressing their personal standards. Tangney et al. (1998), however, argued that self-discrepancies of *all* kinds should result in shame, not guilt, because of shame’s association with global negative self-evaluation, while guilt’s focus is on behaviour (Lewis, 1971). Their results tended to support this notion, and generally failed to provide support for Higgins’s predictions, namely that distinct self-discrepancies are differentially related to shame and guilt. At least three issues, however, are worth pointing out with regard to Tangney et al.’s study.

In the first instance, all self-reported self-discrepancies (e.g., actual/own:should/own, actual/own:ideal/other, etc.) were highly correlated, raising concerns as to whether the scales used were able to measure the distinct constructs that lie at the heart of self-discrepancy theory. In this regard, Tangney et al. (1998) suggested that priming procedures that activate self-discrepancies and thus increase their accessibility, may be more suitable to detect differences between different self-discrepancies (e.g., Strauman, 1992). The current IAT paradigm could certainly be argued to be in line with such an approach, because it was designed to activate self-discrepancies of an actual/own versus should/own nature.

Second, in Tangney et al.'s study, tendencies to experience guilt and shame were measured through use of the TOSCA, which measures guilt and shame *proneness*. As explained in Study 1, guilt and shame proneness are entirely different constructs to frequency-based constructs of guilt and shame. Because Higgins's (1987) postulations predict that self-discrepancies are differentially linked to "chronic tendencies to experience guilt and shame" (Tangney et al. 1998, p. 265), frequency measures, such as the Guilt Inventory, may be more appropriate in evaluating Higgins's claims.

Finally, Tangney et al. (1998)'s overriding, and rather simplifying, account of the role of the self in separating guilt from shame may not be entirely accurate in its interpretation, nor is it the only determining factor that distinguishes these two frequently confused affective states (see Teroni & Deonna, 2008). In particular, H.B. Lewis's (1971) original thesis highlights the different *foci of evaluation* in shame and guilt, rather than just the presence or absence of the self, because the self is implicitly implicated in both emotions. Notably, guilt has intimate connections with responsibility and control, such that one only feels guilty over those actions in which one is somehow implicated (Lamb, 1983). A self-discrepant action can thus result in shame *or* guilt, depending on one's focus of evaluation: If the action is perceived as having a negative import on the global self, shame should ensue; by contrast, guilt should ensue if one views only the action one did in a negative light, without it impacting on the self as a whole (Teroni & Deonna, 2008). Whether shame or guilt ensues is, in fact, where self-discrepancy theory comes into play, i.e., transgressions that undermine one's *ideals* are associated with shame, whereas transgressions that violate a personal or societal *norm* are associated with guilt (Teroni & Deonna, 2008). Tangney et al. (1998)'s claims that all self-discrepant acts should be more closely associated with shame than guilt, is thus unfounded.

Taken together, despite a lack of irrefutable evidence for guilt and shame's relation to distinct self-discrepancies, self-discrepancy theory has become part of mainstream theorising about guilt and shame (Teroni & Deonna, 2008). Moreover, in line with the current study's approach, Tangney et al. (1998) suggested that specific discrepancy-emotion relations may be more readily detected when individuals with large magnitudes of self-discrepancies are selected for participation (cf. Houston, 1990).

Physiological Data

HR data obtained during IAT feedback indicated that the lowest HR reactivity was associated with the Negative IAT condition (i.e., guilt), while the highest HR reactivity was associated with the Positive IAT condition (i.e., pride). Because a counterbalanced order of presentation was followed, these data could not be explained by any order effects; besides, others have failed to find physiological differences due to the order of presentation of neutral, positive, and negative stimuli (Frazier et al., 2004). A more feasible explanation, therefore, is that the magnitude of HR change was determined primarily by the valence of IAT feedback.

HR has a more complicated association with subjective experience than, for example, SCL, which has long been considered one of the most sensitive measures of emotional arousal (Cacioppo & Tassinary, 1990). Although the recall of both positive and negative experiences may lead to increases in HR, which suggests an association with arousal, HR reactivity during visual perception is considered to be primarily determined by valence, such that unpleasant stimuli effect relatively greater HR deceleration (Lang et al., 1993). My finding of a positive association between HR reactivity and the pleasantness of IAT feedback is therefore consistent with previous studies that made use of either emotive pictures or film stimuli to elicit emotion (Bradley, Cuthbert, & Lang, 1990; Britton, Taylor et al., 2006; Frazier et al., 2004; Greenwald, Cook, & Lang, 1989; Lang et al., 1993).

HR responses obtained during the current fMRI paradigm, however, were in sharp contrast to the findings of Study 1, where Guilt condition participants displayed the highest HR reactivity, while Pride condition participants displayed only low cardiovascular arousal. Although the physiological responses of Study 1 and 2 therefore appear to be incongruous, they may be explained by a closer inspection of the two emotion manipulation paradigms that were employed. Moreover, these apparently diverging physiological responses underscore the fact that any emotional response is highly context dependent (Stemmler et al., 2001), making it difficult to compare physiological responses from tasks that differ in their physical contexts, or in the contextual demands of the emotion eliciting event. Differences between the emotion elicitation paradigms of Study 1 and 2 are considered in the *General Discussion*.

Neural Activations During the Negative IAT/Guilt Condition

To verify that neural activations detected during the Negative IAT feedback condition were not confounded by effects of task difficulty during the preceding IAT button-press task, response times to all IAT stimuli were analysed (Chen et al., 2008). Although response times to stimuli in all experimental conditions were fairly similar, participants on average responded faster to stimuli in the Negative IAT stimulus condition. The most likely explanation for this finding is that the Negative IAT sorting condition was easier. While it may also be possible that the Negative IAT condition stimulated enhanced attention and therefore better task performance (to avoid further negative feedback), this explanation is unlikely, because faster response times during Negative IATs were observed even before a feedback pattern could be established. Because Negative IATs could therefore not be viewed as more difficult than Neutral IATs, neural activations observed during Negative IAT feedback could not be attributed to any additional cognitive demands during those periods.

Only one previous neuroimaging study has looked directly at brain activations associated with the transient experience of self-relevant guilt (Shin et al., 2000). Shin and colleagues, however, made use of a script-driven imagery paradigm during which participants recalled a previously experienced guilt episode. Despite significant differences between the emotion elicitation paradigm employed by Shin et al. and the current investigation, remarkable similarities in terms of brain activation patterns are observed: Both studies detected significant neural activation in anterior paralimbic structures, including the supragenual ACC and left anterior insula/ventrolateral PFC. Several areas, in particular the medial prefrontal cortex, however, were additionally recruited during the current guilt manipulation. The current study's prominent mPFC, ACC, and insula activations are also consistent with the notion that impaired processing in these areas contribute significantly to diminished self-conscious emotional responding in patients with frontotemporal lobar degeneration (Sturm et al., 2008).

By comparing the neural activations observed during the current investigation with that of previous moral emotion imaging studies, an important observation is made: The current emotion elicitation paradigm produced significant activations in areas associated with increased emotional arousal (e.g., ACC, insula, thalamus) (Critchley, 2005; Phan, Taylor et al., 2004), while studies that employed emotive sentences or vignettes to elicit moral emotions frequently did not obtain activation in these areas (e.g., Berthoz et al., 2002; Finger et al., 2006; Takahashi et al., 2004, 2008). Another readily apparent observation, compared to previous investigations (e.g., Finger et al., 2006; Kédia et al., 2008; Shin et al., 2000; Völlm et al., 2006; Zahn, Moll et al., 2009), is that temporal lobe activations were not observed *during* the current guilt manipulation, but occurred afterward, during the Negative IAT fixation period. The 20-s fixation periods after IAT feedback may be likened to the post-emotion manipulation periods of Study 1, because they also occurred directly after the emotion manipulation and were of a similar length. The fMRI paradigm could therefore also be employed to detect neural activations after the guilt induction, given that emotional responses may persist after the emotion stimulation period (Garrett & Maddock, 2001).

Frontal lobe and ACC areas. In contrast to the three discrete frontal/paracingulate areas detected in the whole-brain analysis, the area of activation in the reduced cortex volume (i.e., the mask analysis) encompassed one large area of activation within the mPFC. This area extended upward to include dorsomedial PFC, downward to include ventromedial PFC, and posteriorly to include supragenual anterior cingulate cortex (using the nomenclature of Northoff et al., 2006; Figure 19). Rather than making global inferences about the nature of this large mPFC activation, however, the three distinct activations observed during the whole-brain analysis, namely the superior frontal gyrus/supraACC (-4,28,30)[9], superior frontal gyrus (-4, 43,18), and pregenual ACC (-7,34,15), will be considered in more detail because they pertain to more circumscribed anatomical areas within the frontal lobes.

---

[9]All coordinates referred to in this thesis are based on the Talairach system of coordinates.

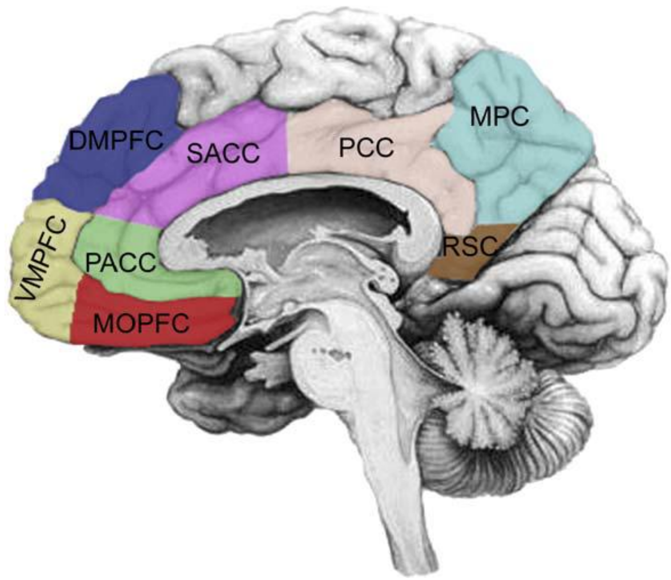


Figure 19. Schematic illustration of cortical midline structures. MOFC: medial orbital prefrontal cortex (BA 11, 12), VMPFC: ventromedial prefrontal cortex (BA 10, 11), PACC: pre- and subgenual anterior cingulate cortex (BA 24, 25, 32), SACC: supragenual anterior cingulate cortex (BA 24, 32), DMPFC: dorsomedial prefrontal cortex (BA 9), MPC: medial parietal cortex (BA 7,31), PCC: posterior cingulate cortex (BA 23), RSC: retrosplenial cortex (BA 26, 29, 30) (adapted from Northoff et al., 2006).

The first, more dorsal, superior frontal gyrus/dorsomedial PFC activation (BA 8/9) detected in the present investigation corresponds to the more posterior, cognitive area of rostral medial frontal cortex (Amodio & Frith, 2006). Dorsomedial together with ventrolateral PFC (BA 47), which was also activated during the present Negative IAT/Guilt condition, have been observed in several studies investigating neural responses to socially inappropriate or embarrassing scenarios (e.g., Berthoz et al., 2002; Takahashi et al., 2004; Zahn, Moll et al., 2009). Finger et al. (2006) extended previous accounts by suggesting that activation in these areas pertain to the behavioural response demands of an event, rather than the specific event per se. They assessed neural activations associated with moral and embarrassing/social transgressions in the presence or absence of an audience. Results showed increased activity in left dorsomedial and ventrolateral PFC *only* when there was a need for behavioural change, thus social/embarrassing transgressions with no witnesses to the event did not activate these areas because no behavioural responses were necessary in those situations. In line with guilt's behavioural motivation to bring about restitution or appeasement actions (Lindsay-Hartz, 1984; Tangney et al., 1996), Finger et al. (2006) interpreted enhanced activity in dorsomedial and ventrolateral PFC as signaling inappropriate or unacceptable social behaviour, which is thought to result in the initiation of alternative motor responses. In this way, behaviour is changed away from that which prompted the aversive response, to a more beneficial option.

The above interpretation of the role of dorsomedial and ventrolateral PFC activation during moral transgressions is consistent with previous imaging work on more basic response reversal paradigms: Neural activation associated with a change in response also typically includes the dorsomedial and ventrolateral PFC (Cools, Clark, Owen, & Robbins, 2002; O'Doherty, Critchley, Deichmann, & Dolan, 2003; Remijnse, Nielen, Uylings, & Veltman, 2005). These areas are, however, not only active during cognitive tasks of changing reinforcement contingencies (O'Doherty, Kringelbach, Rolls, Hornak, & Andrews, 2001; Rolls, 2000), but are also active in response to social cues that modify current behavioural responses (e.g., Kringelbach & Rolls, 2003; Shafritz, Collins, & Blumberg, 2006). To explain the profound social interaction problems of neurological patients with OFC damage, Blair and Cipolotti (2000) formulated the social response reversal model. According to this model, OFC patients may have reduced ability to activate representations of situations that have previously been associated with negatively-valenced reactions from others, or these representations can no longer modify their current behaviour in situations where it is required. Their patient with bilateral OFC damage was impaired at identifying social transgressions, as well as at altering his behaviour in response to socially aversive cues, such as angry and disgusted facial expressions.

The distinct connectivity and function of ventrolateral PFC in diverse imaging paradigms lend further support to the notion that aversive moral and social emotions may serve as social cues that generate alternative motor responses. Notably, lateral compared to medial OFC receives more multimodal sensory-related inputs (Carmichael & Price, 1996), and appears to be more involved in changing responses under unexpected circumstances (Elliott, Dolan, & Frith, 2000). Lateral OFC is also commonly activated by negative emotional expressions (Basile et al., 2011; Blair, Morris, Frith, Perrett, & Dolan, 1999; Sprengelmeyer, Rausch, Eysel, & Przuntek, 1998), incongruent IAT trials (Chee, Sriram, Soon, & Lee, 2000; Luo et al., 2006), and importantly, is activated in response to a wide range of punishing stimuli (Kringelbach & Rolls, 2004; O'Doherty et al., 2001; Remijnse et al., 2005; Seymour, Singer, & Dolan, 2007).

Ventrolateral PFC therefore appears sensitive to cues of punishment or negative emotional reactions from others, which signal to us that our current behaviour is socially unacceptable and should be curtailed or changed (Blair & Cipolotti, 2000; Kringelbach & Rolls, 2004).

Taken together, dorsomedial and ventrolateral PFC activation during guilt may be responsible for processing specific social, contextual, and emotional cues to modify current behaviour (Finger et al., 2006). This interpretation is compatible with Moll and colleagues' view of the PFC as being involved in the context-dependent representation of events during moral cognition (Moll, Zahn et al., 2005), as well as with evidence from disorders affecting the frontal lobes, where affected patients typically act in ways that are not socially acceptable (Eslinger, Flaherty-Craig, & Benton, 2004; Miller, Darby, Benson, Cummings, & Miller, 1997; Tranel, 1994).

The second superior frontal gyrus activation (BA 9/32; -4, 43, 18) falls within the more anterior part of rostral frontal cortex, which has been associated with emotional tasks (Amodio & Frith, 2006). This activation, in fact, has very similar peak coordinates to the area suggested by Steele and Lawrie (2004) to be concerned with emotion (mean Talaraich coordinates ±5, 46, 18), based on a comprehensive meta-analysis of emotion-inducing and cognitive task imaging studies. It also corresponds closely to the area of maximum activity detected during 'real-time' mentalising tasks, namely the anterior paracingulate cortex bilaterally (-10, 50, 30 and 8, 54, 12) (Gallagher & Frith, 2003; Gallagher, Jack, Roepstorff, & Frith, 2002). A similar area (-8, 42, 20) has also been implicated in inferring the communicative intentions of another, for example, judging someone as being deceptive (Grèzes, Frith, & Passingham, 2004).

Anterior rostral frontal cortex (BA9/32) therefore appears intimately involved in emotion induction on the one hand, but also in more abstract mentalising processes, on the other hand. Amodio and Frith (2006) integrated these accounts by arguing that this area is not necessarily involved in intrinsic emotional processing *per se*, but rather plays a crucial role in our ability to reflect on how we (and others) perceive ourselves, particularly with regard to 'hot'/emotional mental states. Supporting this explanation, Ochsner et al. (2004) found activation within anterior rostral frontal cortex to be associated with assessing one's own emotional state, which is very similar to assessing self-knowledge. Imaging studies of self-referential tasks have also consistently pointed to the involvement of mesial aspects of BA 8, 9, and 10 for spontaneous as well as task-related introspective mental activity (Gusnard, Akbudak, Shulman, & Raichle, 2001;

Northoff et al., 2006). Finally, Johnson (2002) detected anterior mPFC activity specifically in relation to self-reflective thought, in a task where participants were required to evaluate their own abilities, traits, and attitudes. I therefore propose that activity in the anterior rostral frontal cortex during the current Negative IAT/Guilt condition stemmed from participants' engagement in self-reflection.

The first ACC activation reported in the present study, i.e., supragenual ACC (BA 32) (using the nomenclature of Northoff et al., 2006), is located approximately 30 mm above the anterior commissure–posterior commissure (AC–PC) plane, directly above the genu of the corpus callosum. This area is situated just posterior to Steele and Lawrie's (2004) boundary that separates the cognitive (dorsal) and emotional (ventral) divisions of the ACC. The current supraACC activation may therefore be described as more cognitive in function (Bush, Luu, & Posner, 2000). The second ACC activation is located 15 mm above the AC-PC plane, slightly dorsal to the subgenual cingulate, and directly anterior to the genu of the corpus callosum. It therefore corresponds to pregenual cingulate cortex (pACC) (Johansen-Berg et al., 2008), and may be associated more with emotional processing (Amodio & Frith, 2006).

The more dorsal ACC activation (supraACC, BA 32) observed in the present study has also previously been detected in association with moral transgressions, moral emotions, and moral judgment tasks (e.g., Greene et al., 2004; Immordino-Yang et al., 2009; Kédia et al., 2008; Shin et al., 2000). Moreover, in moral judgment tasks, enhanced ACC activity is associated with longer reaction times and increased decision difficulty in response to complex moral dilemmas (Greene et al., 2004), while it does not activate in response to simple ethical decision-making (Heekeren et al., 2003). ACC activity also appears to be modulated by emotional valence: It has been associated with the regulation of negative emotional stimuli (Mak, Hu, Zhang, Xiao, & Lee, 2009a; Ochsner, Ray et al., 2004), and is generally more associated with negative than positive emotional experiences (George et al., 1995; Prohovnik et al., 2004). Notably, dorsal ACC may serve to direct attention toward subjective emotional states (Lane et al., 1998; Phan, Liberzon, Welsh, Britton, & Taylor, 2003), and is more consistently recruited in emotion tasks that also involve cognitive components (Phan et al., 2002).

Together, these findings support dorsal/supragenual ACC's important role in cognitive control, and more specifically, the detection and monitoring of conflicts (Botvinick et al., 2004; MacDonald, Cohen, Stenger, & Carter, 2000). The ACC is considered to play a central role in cognition by continually monitoring response tendencies for competition, and signaling the need for enhanced top-down control when conflicts are detected (van Veen et al., 2001). In the Eriksen flankers task, for example, increased conflict between intention and behaviour on incompatible flanker trials are associated with response errors (Gehring, Goss, Coles, Meyer, & Donchin, 1993). The event-related potential (ERP) associated with such errors, the error-related negativity (ERN), has been localized to the ACC (Dehaene, Posner, & Tucker, 1994) – a finding supported by computational models (Botvinick, Braver, Barch, Carter, & Cohen, 2001), as well as neuroimaging studies of the flanker task (Casey et al., 2000; Durston et al., 2003). Amodio et al. (2004) have shown that these conflict-detection processes are also sensitive to the automatic, but undesired, activation of racial stereotypes in low-prejudice individuals.

In the present study, the prominent dorsal ACC activation may thus correspond to internal conflict experienced by participants, i.e., between their internalized standards and actual prejudiced IAT feedback (see, e.g., Holroyd et al., 2004). This explanation fits well with a finding from Study 1, namely that guilt was significantly correlated with BIS sensitivity: Amodio and colleagues (2008) recently found BIS (as measured by the Carver and White BIS/BAS scales) to be associated with ACC-related conflict monitoring.

Finally, another function associated with ACC activity is the generation of increased autonomic output, e.g., cardiovascular arousal (Critchley et al., 2003). This hypothesis is considered in more detail in the *General Discussion.* It is worth mentioning here, however, that the autonomic control account of ACC function is not necessarily distinct from its conflict-monitoring role: Increased autonomic arousal is a robust feature of increasing cognitive effort. Moreover, conflict detection may be coupled with increased negative affect, which is also associated with heightened autonomic arousal (McNaughton & Corr, 2004).

Although activations discussed thus far have corresponded to fairly well-characterized regions with regard to their functional specialization, the functional significance of pACC (BA 24/32) activity in the present guilt manipulation is more debatable. Of interest is the fact that the pACC area activated in the present investigation is associated with areas involved in a default, self-monitoring, state of brain activation. More specifically, pACC lies within ventral mPFC, which typically exhibits decreased activity from baseline during more attention-demanding, goal-directed behaviours (Gusnard et al., 2001; Raichle et al., 2001). This decrease, however, has been found to be less pronounced when emotion processing co-occurs with the attention-demanding task (Simpson, Snyder, Gusnard, & Raichle, 2001), and may even appear as an increase if an attention-demanding control task is used as baseline, as in the present situation. Gusnard et al. (2001) thus hypothesized that ventral mPFC is involved in assessing the salience of emotional stimuli.

In line with an emotion-processing account, pACC activity in the present investigation was intimately related to participants' subjective emotional experiences, such that activity in this area had a significant negative linear relationship with self-reports of guilt. Increased pACC activity thus appeared to inhibit participants' affective experiences. Consistent with this interpretation, a region slightly inferior to the present pACC activation has previously showed enhanced activity when participants were required to *decrease* their affective responses to negative emotional stimuli by using a self-focused strategy (Ochsner, Ray et al., 2004). Likewise, in an emotional Go/No-Go task, inhibition for sad faces in the No-Go condition preferentially activated pregenual ACC (BA 24/32) (Shafritz et al., 2006). Shafritz et al. (2006) reasoned that the unique location of pACC, i.e., at the intersection of dorsal and ventral ACC subdivisions (Steele & Lawrie, 2004), placed it in an ideal position to integrate cognitive and emotional processes. Perigenual ACC's extensive connections to limbic, as well as frontopolar and dorsal cingulate areas, support this interpretation (Johansen-Berg et al., 2008), and suggest that it may be involved in integrating visceromotor aspects of emotional processing with cognitive information.

In the present context, it may therefore be that (i) participants actively tried to down-regulate feelings of guilt, perhaps to avoid it from influencing their performance on subsequent IATs; and that (ii) those who were better at this emotion regulation, experienced less affective distress. In support of this conclusion, areas within the medial and ventromedial frontal cortex have been associated with regulating limbic emotion structures (e.g., the amygdala), such that enhanced activity in these areas lead to the suppression of limbic structures (Phelps & LeDoux, 2005). Posttraumatic stress disorder (PTSD) patients, for example, have been associated with reduced activation in medial prefrontal areas during exposure to traumatic memories, compared to control participants (Bremner et al., 1999). In addition, Rauch et al. (2003) detected decreased volumes of pregenual ACC in PTSD patients compared to controls. The present results thus provide support for the hypothesis that medial prefrontal dysfunction contributes to the inability to suppress bad thoughts and memories in patients with affective disorders (cf. Shin et al., 2004).

Finally, because stimuli employed in negative emotion manipulation paradigms may also be conceived of as "painful stimuli" (Amodio & Frith, 2006, p. 274), a different way of understanding pACC involvement during guilt elicitation may relate to the experience of pain. By this notion, increased pACC activity may correspond to the 'blocking out' of affective pain. This interpretation is considered in more detail in the *General Discussion*.

Before moving away from the frontal lobes to consider more posterior brain activations, it is of interest to observe that no neural activation was detected in the subgenual cingulate cortex (SCC) during guilt. While the SCC has consistently been associated with reliving sad memories (Liotti et al., 2000; Phan et al., 2002), it has more recently also been associated with feelings of guilt when participants imagined performing moral transgressions against a known other (Zahn, Moll et al., 2009). In a subsequent study, Zahn and colleagues refined their interpretation of activity in this area and suggested that SCC activation reflects individual differences in empathic responding, which is an integral component of various moral sentiments (e.g., guilt and compassion) (Zahn, de Oliveira-Souza et al., 2009). Consistent with this interpretation, SCC activity has been observed in association with feelings of romantic and maternal love (Bartels & Zeki, 2004), as well as during charitable, altruistic decisions to donate money, presumably driven by affiliative feelings toward a societal cause (Moll et al., 2006). Zahn and colleagues thus argued that activity in this region is driven by attachment-related experiences (Insel & Young, 2001; Zahn, de Oliveira-Souza et al., 2009).

The negative finding for SCC activity in the present investigation may be explained by susceptibility artifacts often problematic in ventral frontal areas. Alternatively, if the SCC was activated equally in the Neutral IAT condition, it may not have showed up after subtraction. In this regard, Moll and colleagues (2007) have shown that neutral agency robustly engages ventral and subgenual PFC sectors. A more likely explanation, however, is that the lack of SCC activation in the present study is a consequence of the nature of the guilt-eliciting paradigm. The paradigms employed by Zahn and colleagues consisted of scenarios describing how you personally harm someone familiar to you, or act in a way counter to established social norms toward that person (e.g., your mother) (Zahn, de Oliveira-Souza et al., 2009; Zahn, Moll et al., 2009). The present paradigm, however, involved the transgression of a personal norm without any interpersonal harm. It is likely that areas associated with attachment (e.g., the SCC) may only be activated during guilt when another individual is harmed and one feels compassion for the victim. In line with this reasoning, Basile et al. (2011) argued that the SCC is more likely to be associated with altruistic than deontological guilt, while deontological guilt in their study was associated with a more dorsal ACC activation (-4, 30, 24). Because guilt evoked in the present study was more in line with deontological than altruistic guilt, my results fully support the distinction proposed by Basile et al. (2011).

Posterior and subcortical/limbic areas. While the anterior cingulate cortex is "executive" in function, the posterior cingulate cortex has been described as "evaluative" (Vogt, Finch, & Olson, 1992), because it is associated with the evaluation and monitoring of sensory information (Raichle et al., 2001). In terms of emotional processing, the posterior cingulate/retrosplenial cortex is one of the most consistently activated areas in neuroimaging studies of emotion, particularly in response to emotionally salient stimuli (Basile et al., 2011; Britton, Phan et al., 2006; Harenski & Hamann, 2006; Maddock, 1999). Its prominent role in emotion, together with the hippocampus, appears to involve autobiographical emotional recall. For example, the posterior cingulate activates with self-generated emotions and listening to autobiographical memory scripts (Damasio et al., 2000; Fink et al., 1996). It may therefore link emotion and episodic memory processes (Maddock, 1999). The posterior cingulate is also commonly implicated in ToM (Fletcher, Happe et al., 1995), as well as in moral judgment (Greene et al., 2004).

Consistent with the interpretation that the posterior cingulate activation reflects episodic memory recall, I also observed significant activation in the right hippocampus during guilt. It has been suggested that the posterior cingulate and retrosplenial cortices may serve to connect the dorsolateral PFC with mesial temporal areas associated with memory, e.g., the hippocampus (Goldman-Rakic, Selemon, & Schwartz, 1984). The hippocampus is implicated in the processing of autobiographical memory, and may facilitate the conscious retrieval of salient memories that allow previous experiences to influence current behaviour (Casebeer, 2003). Posterior cingulate activation was detected in Britton et al.'s (2006) imaging study in response to social, but not non-social, film clips. Interestingly, participants in that study reported that some of the emotive film clips triggered personal memories associated with the footage. Likewise, in the current guilt paradigm, it is possible that participants engaged in self-reflection and remembered personal situations where they have responded with prejudice. Recall of emotional memories may thus account for the posterior cingulate and hippocampus activations observed in the present study.

The neighbouring precuneus area is another posterior area that has been implicated in moral cognition and ToM (Greene & Haidt, 2002; Moll, Zahn et al., 2005). For example, in Kedia et al. (2008)'s study of moral emotions that result when "an agent harms a victim", the precuneus was consistently activated in scenarios that also involved someone else, including guilt, other-anger, and compassion, but not during scenarios where the self acted as both the agent and victim (i.e., self-anger). Like other areas associated with mentalizing, the precuneus is likely to have a distinct role in the process of perceiving and reasoning about others (Saxe, 2006). Together with the posterior cingulate, it appears to be involved in memory, specifically affective imagery, in the context of coherent social narratives (Fletcher, Frith et al., 1995). The significant positive correlation between blood flow in the posterior cingulate and precuneus areas supports this conclusion and suggests that these regions may be functionally connected.

An important finding of the present study was that the guilt condition was associated with significant insular activity: both the left anterior- and right posterior insular cortices showed significant activation. The insula is a structure intimately associated with emotional processing. Its function as visceral sensory area may serve to represent and evaluate internal feeling states of the organism (Craig, 2002; Damasio et al., 2000). The anterior insula is also strongly implicated in empathic responding (Lamm, Decety, & Singer, 2011; Singer et al., 2004). In this regard, empathy is believed to automatically evoke a similar feeling state in the observer as in the individual being empathized with (de Vignemont & Singer, 2006; Gallese, Keysers, & Rizzolatti, 2004).

Consistent with the notion that participants engaged in self-reflection during guilt, Reiman et al. (1997) observed that insula activation is more readily associated with evaluative and experiential aspects of self-generated, rather than externally-generated, emotions. This observation was supported in a subsequent meta-analysis of emotion induction studies (Phan et al., 2002). Moreover, Phan et al. (2004) reported increased anterior insula activity in connection with increased emotional intensity (regardless of valence), as well as increased personal association with experimental stimuli (see also Kircher et al., 2000). With regard to more specific affective states, anterior insula activation has been associated with negative emotions, e.g., sadness and guilt (Basile et al., 2011; George et al., 1996; Shin et al., 2000), as well as aversive stimuli, such as tasting salt (Kinomura et al., 1994). In particular, feeling disgusted or perceiving disgust in others is strongly associated with activation in the anterior insula

(Fitzgerald et al., 2004; Phillips et al., 1997; Sprengelmeyer et al., 1998; Wicker et al., 2003). Taken together, the above accounts underscore the fact that participants were emotionally aroused during the guilt manipulation, and found the material self-relevant.

The anterior and posterior portions of the insula differ considerably in terms of their cellular organization and functional connectivity (Mesulam & Mufson, 1982a, 1982b). Whereas the anterior insula is agranular and extensively connected to anterior paralimbic and subcortical emotion structures (e.g., the amygdala), the smaller posterior portion is granular and shares connections with superior temporal, parietal, premotor, and somatosensory cortices (Augustine, 1996). The posterior portion therefore appears to be more involved in processing somatosensory and visceral information from the body, e.g., the experience of physical pain (Peltz et al., 2011; Singer et al., 2004).

The precise functional significance of the right posterior insula activation during the current guilt condition, and, in particular, its negative association with subjective reports of guilt, however, is unclear. Given the involvement of the right inferior posterior insula in cardiovascular control, it may be possible that activation in this area served to adjust autonomic outflow during guilt (Williamson, Fadel, & Mitchell, 2006). This interpretation is considered in more detail in the *General Discussion.*

The thalamus can be described as a central sensory gateway, given its role in relaying bodily afferents to higher cortical and subcortical areas (Craig, 2002). Thalamic activation is readily observed in emotion induction studies, irrespective of valence or the precise method of induction (Britton, Phan et al., 2006; Damasio et al., 2000; Lane, Reiman et al., 1997; Moll, de Oliveira-Souza, Eslinger et al., 2002; Reiman et al., 1997). It was therefore not surprising to detect activation in this structure during guilt. The negative finding for amygdala activation in the present study, however, was more puzzling. While this finding appeared at first to run counter to my predictions for guilt, it may be accounted for by several explanations. These are considered in the *General Discussion*.

Neural activation during the fixation period. I observed significant neural activation in several areas during the Negative fixation period compared to the Neutral fixation period. These activations were observed despite the fact that participants fixated on a similar cross in both conditions, and strongly suggest that the effects of the guilt manipulation extended past the emotion induction period. Notably, several temporal lobe areas (in addition to the mPFC, ventrolateral PFC, and posterior cingulate) showed significant activation, which were not observed during the emotion elicitation period. These areas included the bilateral temporal poles, left mesial and lateral middle temporal gyrus, and the bilateral posterior STS/TPJ. Neural activations in the mPFC, bilateral posterior STS/TPJ, and anterior temporal poles support the notion that participants engaged in (implicit) mentalizing during the Negative fixation condition (Frith & Frith, 1999; Gallagher & Frith, 2003; Saxe & Kanwisher, 2003).

The importance of the temporal lobes in social cognition has been demonstrated in various neuroimaging studies, as well as in patients with specific neurological impairment (Frith & Frith, 2003; Olson, Plotzker, & Ezzyat, 2007; Zahn et al., 2007). Moll and colleagues (2005) argued that temporal areas are responsible for storing and extracting socially significant perceptual and semantic information from the environment. In particular, they consider the posterior STS a key region for storing social perceptual features, which may be gained by extracting various social cues from the environment, e.g., facial expressions, eye gaze, and body language (Allison et al., 2000; Boddaert et al., 2004; Puce, Allison, Bentin, Gore, & McCarthy, 1998). The posterior STS therefore appears to be active when observing living things, but importantly, is also active during the retrieval of previously acquired information about living things (Frith & Frith, 2003).

Compared to the posterior STS, the anterior temporal lobes are thought to store social functional features, which may be defined as context-independent semantic properties of social situations (Moll, Zahn et al., 2005). Zahn and colleagues (2007) identified the right anterior temporal lobe (aTL, BA38/22) as a neural substrate specialized for the representation of abstract functional (i.e., nonsensory) semantic knowledge. This area thus enables us to grasp the meaning of social concepts, such as loyalty, honour, tactlessness, and by extension a person's social behaviour, without sensory detail. The importance of the anterior temporal lobes in storing such abstract social semantic representations is underscored by impairments in conceptual knowledge and social behaviour in semantic dementia (Bozeat, Gregory, Ralph, & Hodges, 2000; Davies et al., 2005), and in patients with anterior temporal lobe resections (Lu et al., 2002).

In light of the above accounts of the functional significance of anterior and posterior temporal areas in social cognition, I expected to find activation in these areas during the guilt manipulation. The most feasible explanation for the lack of significant temporal activation during guilt, however, is that temporal areas were activated equally during the Negative and Neutral IAT feedback periods, and may thus have been cancelled out during subtraction. This explanation seems plausible, given that the Neutral IAT condition involved a similar amount of social contextual information as the Negative IAT condition.

Alternatively, participants may have only engaged in mentalising during the fixation period. They may, for example, have reflected on the experimenter's opinion of their performance. Interestingly, the current Negative fixation condition bore a close resemblance to the embarrassment condition in the Takahashi et al. (2004) study, with overlapping areas of activation in several frontal and temporal regions. As mentioned previously, it may be possible that participants experienced heightened embarrassment because they were concerned about others' evaluation of their performance (Finger et al., 2006). Nevertheless, increased activation in the posterior cingulate, together with other well-established ToM areas, support the view that participants engaged in continued self-reflection and mentalising during the post-manipulation period (Johnson et al., 2002).

Behavioural Motivation

In contrast to findings from Study 1, self-reports of guilt in the current fMRI study were not significantly correlated with BIS sensitivity, but showed only a weak positive relation. Higher BIS scores, however, were significantly associated with decreased activity in the pACC and right posterior insula. By comparison, correlations between BAS sensitivity and guilt showed a strong inverse relationship, such that lower BAS scores were associated with higher self-reported guilt. BAS scores were also positively associated with signal increases in the pACC and right posterior insula.

While the lack of association between BIS and guilt may appear to contradict the findings of Study 1, they may be explained when one considers the original selection criteria. In Stage I of the fMRI study, individuals with high BIS sensitivity were explicitly chosen for further participation above those with low BIS scores. This selection criteria resulted in a relatively narrow range of BIS scores across participants at the higher end of the scale's distribution (*Range* = 10, *SD* = 2.76, compared to Study 1: *Range* = 14, *SD* = 3.88). By comparison, because selection criteria was not specific for BAS sensitivity, BAS scores were much more distributed (*Range* = 17, *SD* = 3.67), and may therefore have had a bigger effect on the levels of experimentally-induced guilt than BIS scores. A more detailed discussion of guilt's behavioural motivation is presented in the *General Discussion*.

Limitations

The current fMRI study consisted of a carefully designed paradigm, based on psychological theory as well as data from a pilot study, to ensure effective and specific moral emotion elicitation. The experimental conditions were matched in terms of the valence and arousal values of IAT stimuli (Appendix I), as well as the complexity of IAT feedback sentences (Appendix J), to ensure that task-related effects did not interfere with emotion activation results. A number of study limitations, however, should be considered.

The biggest limitation of the present investigation was the lack of power of the paradigm to discriminate successfully between the Positive and Neutral IAT conditions. While the pride manipulation appeared successful, based on subjective emotion ratings, no significant neural activation was detected in the Positive IAT condition. It may be possible that the feedback period over which emotion was measured was too long to detect significant brain activation associated with pride. In this regard, a recent fMRI study reported that participants found it easier and were more effective at regulating positive than negative emotion (Mak et al., 2009a). Alternatively, it may be that the emotion elicited in the Positive IAT condition was either not intense enough, or not sufficiently different from the Neutral condition, to detect significant neural activation. This lack of differentiation was also evident in the physiological data analysis: HR reactivity during the Positive IAT condition was not significantly different from that in the Neutral IAT condition.

Because the present study failed to detect significant brain activation associated with pride, it should be noted that some previous imaging studies also failed to detect significant neural activation, or detected only minimal brain activation, in association with positive affect (see, e.g., George et al., 1995, 1996; Takahashi et al., 2008). Positive emotion, in general, therefore appears to be associated with less brain activation than negative emotion, or is less intense when evoked by typical neuroimaging manipulations. To enhance the statistical power of neuroimaging analyses, it may be beneficial to search within predetermined regions of interest when investigating neural areas associated with positive affect (e.g., Zahn, Moll et al., 2009).

Technical limitations in the current study included the limited physiological data that I was able to capture during the fMRI paradigm. Because of a lack of available fMRI-compatible devices, I was not able to record electrodermal activity. Furthermore, because of the relatively short emotion elicitation periods (30 s), I was only able to calculate HR reactivity, and not indices of HRV (Task Force, 1996). Additional physiological data would have enabled a richer interpretation of significant brain activations observed during guilt, as well as more comprehensive comparisons with physiological data from Study 1.

Another technical limitation was the difficulty in detecting neural activation in anterior temporal and orbitofrontal areas, due to signal dropout in these areas. Although Shin et al. (2000) did not detect OFC activation in their PET study of guilt either, techniques tailored to optimize the detection of signal in areas associated with magnetic susceptibility artifact, e.g., z-shimming, should perhaps be explored in future (Glover, 1999). Such techniques are, however, often associated with other limitations, e.g., temporal resolution (Zald, 2003).

A common confounding factor of moral emotion elicitation studies is that the purity of the emotional state is often compromised, because the paradigm may also evoke other (basic) emotions in parallel (Takahashi et al., 2004). Activations in the current Negative IAT condition may thus have reflected changes in other emotions in addition to guilt. Although the Negative IAT condition was marked by high self-ratings of guilt, several other negative emotions, including embarrassment, shame, anxiety, and anger were also elevated in this condition. In this regard, anxiety has also previously been associated with paralimbic activation (Chua, Krams, Toni, Passingham, & Dolan, 1999; Kimbrell et al., 1999), while anger has been associated with left OFC/ventrolateral PFC activation (Dougherty et al., 1999; Kimbrell et al., 1999). As argued in Study 1, however, these emotions may form part of the emotional profile of guilt, and may only be separated out from activations associated with "pure" guilt in an experiment where these emotions are elicited separately.

Finally, a common limitation of all fMRI studies of emotion, yet worth mentioning again, is the influence that the artificial nature of the MRI surroundings may have on effective emotion elicitation. Lying on one's back while focusing on not moving, and with loud hammering noises in the background, is bound to impact on the nature of the emotion elicited. These particular confounds, in addition to technical constraints in the kinds of stimuli that can be used, and difficulties in staging ecologically valid paradigms within the MRI chamber, will continue to beleaguer future fMRI studies of emotion. Continued and innovative efforts at overcoming these difficulties will be necessary if we are to learn how emotion is instantiated in the brain.

GENERAL DISCUSSION

This thesis was designed with the aim of empirically disentangling the physiological and neural substrates of guilt in order to inform our theoretical understanding of this complex moral emotion. To accomplish this goal, I developed two novel paradigms for the real time elicitation of guilt, with pride serving as a contrasting positive moral emotion. Both paradigms were devised to be emotionally arousing, self-relevant, and in particular, afford participants with agency. Personal agency was deemed especially important given the fact that guilt (but not shame) is considered to have intimate links with responsibility and control (Lamb, 1983). Both paradigms also necessitated a degree of deception in order to elicit natural and intense emotional responses that would be unbiased by desirability concerns.

The emotion elicitation paradigm in Study 1 consisted of a social psychology experiment imbued with high ecological validity, in that the laboratory scenario was representative of the kind of psychological stressors individuals might encounter in real life (Cacioppo et al., 1994). This approach is of great value because results are more predictive of cardiac reactivity patterns in response to real emotion-inducing situations, and may inform current theorizing on the impact of guilt and pride on our psychological well-being. Guilt was elicited based on the perception that one has done something "bad", i.e., a specific act (taking money that does not belong to you), that impacted negatively on someone else (the research assistant was dismissed) (Lindsay-Hartz, 1984).

The fMRI emotion elicitation paradigm employed in Study 2 was based on findings from literature on prejudice and self-discrepancy theory: Guilt was elicited in low-prejudice individuals by subjecting them to a task in which they received discrepant implicit prejudice feedback about themselves. A relatively stringent screening procedure helped to identify suitable participants for this study: individuals with strong internalised, nonprejudiced, *ought* standards (high IMS), but with naturally occurring large discrepancies between their *actual* and *ought* responses (high EMS), who have been shown to be likely to experience strong guilt (Higgins, 1999; Higgins et al., 1986; Plant & Devine, 1998). In addition, high BIS-sensitive individuals were selected based on data from a pilot study, as well as Study 1.

Because the fMRI paradigm satisfied most conditions that have been proposed to increase the likelihood of specific discrepancy-affect relations, e.g., a large magnitude of self-discrepancy, enhanced accessibility of the self-discrepancy, and significant importance of the self-discrepancy to the person possessing it (Higgins, 1999), there was sufficient reason to believe that the paradigm would be effective in eliciting specific emotions of guilt. In addition, the nature of the “transgressions” in the Negative IAT condition was more consistent with failures of prescriptive than proscriptive moral regulation, which has recently been associated with guilt (Sheikh & Janoff-Bulman, 2010). Prescriptive morality’s focus is on what we *should* do rather than on what we *should not* do, and is therefore associated with prosocial behaviour and the presence of a desired goal (Janoff-Bulman et al., 2009). In the current context, participants’ desired goal probably would have been to possess a truly low-prejudice attitude toward different social groups, and to act accordingly. This assumption was supported by self-reports obtained after the emotion manipulation: participants reported wanting to change their behaviours subsequent to the experiment to achieve their personal goal of being nonprejudiced.

Subjective Emotional Experience

Emotional reports obtained from participants in the current investigation provided valuable insights into the subjective experience of guilt. Consistent with past research (Izard, 1991), guilt was associated with various other negative emotions (e.g., disgust, anger, anxiety, and shame), which are likely to be significant motivations in the guilt situation. The particular emotion profile associated with guilt, however, is liable to change from one situation to the next. For example, in the Shin et al. (2000) study, where participants were required to recall a personal episode of guilt, higher levels of sadness were reported compared to findings from Study 1. By comparison, the experimental situation in Study 1 may have led to augmented levels of anxiety and disgust: participants possibly experienced frustration at being unable to repair the situation immediately. In this regard, both Tangney et al. (1992) and Izard (1991) have commented on the potential agony one may experience when reparation (after wrongdoing) is blocked for one reason or another. This increased agony is thought to result from cognitive preoccupation and rehearsal of the guilt-inducing situation.

Regarding the fMRI study, participants also experienced heightened embarrassment and shame in addition to guilt. This co-experience, as mentioned above, may have been due to participants’ awareness of others monitoring their performance (Finger et al., 2006; Smith et al., 2002), despite the fact that the public nature of their performance was not emphasized to them. Subjective emotion reports also indicated, however, that high levels of anger were experienced during the Negative IAT condition. During post-experimental interviews, a number of participants qualified this anger in terms of feeling angry at themselves. Similar findings have been reported by Kédia et al. (2008): their participants rated self-anger nearly as high as guilt in response to guilt-eliciting scenarios. These findings may be explained by the accepted view that guilt is a mixed feeling of self-anger and compassion, but cannot be reduced to one or the other (Baumeister et al., 1994; Ellsworth & Tong, 2006).

Finally, an interesting finding of the present research was that guilt had a significant negative linear relationship with pride in both studies: as guilt increased, pride decreased. One may presume that ameliorative behaviours associated with guilt also serve to restore one's pride and self-esteem. Some evidence for the reverse relationship was also found in Study 1: Pride participants reported significantly reduced guilt from pre- to post-manipulation, which demonstrates the positive effects of pride on psychological well-being.

While subjective reports from both experiments supported the notion that participants experienced guilt rather than shame, the case for 'pure' guilt resulting from a social transgression is questionable. In moral transgression, for example, a person may feel guilty for violating a social standard while at the same time feeling shameful about his/her own shortcomings. Based on subjective data and post-experimental interviews obtained in the present investigation, I believe that guilt is often not as distinct from shame as put forth by June Tangney and colleagues (1996, 2007), and that some element of shame may often be present in a prototypical guilt response. The intensity of this shame, and whether or not the guilt response becomes maladaptive, however, may vary from person to person (see, e.g., Eisenberg, 2000; Orth et al., 2006).

Lewis (1971) argued that guilt and shame are difficult to tease apart when a situation elicits both emotions, because in such a situation the two states tend to blend together and are then typically labeled as "guilt." Others have argued that guilt and shame probably do not co-occur, but may be experienced sequentially or in close contiguity. Clinicians, for example, often observe patients "flip flopping" between guilt and shame (Kubany & Watson, 2003). Guilt and shame may also be difficult to separate because guilt- and shame-related cognitions may be held simultaneously, although it is only those beliefs that are activated in consciousness, with accompanying negative affect, that define the active emotional state (Kubany & Watson, 2003).

Perhaps a confounding factor in distinguishing guilt from shame is lay-people's tendency to say "guilt" when they mean "shame" in everyday situations (Tangney & Dearing, 2002). Guilt may thus be perceived to be a more positive, or desirable, emotion than shame. Unfortunately, researchers are entirely dependent on subjective emotion reports, which are only as accurate as the participant's understanding, and disclosure, of his/her own feelings. In light of these considerations, it is not possible to state unconditionally that guilt was the overriding emotional response in the present investigations. Great care was taken during post-experimental interviews, however, to distinguish between guilt and shame, based on theoretical formulations.

Post-experimental interviews from Study 1 also suggested that the duration of the guilt response varied significantly between individuals: Whereas some participants reported that they rationalized their behaviour straightaway and concentrated on the remaining tasks, others reported feeling "confused," "distracted," and "bad" until the end of the experiment. These differences may reflect emotion regulatory attempts of participants, which are likely to be enhanced in the presence of another or in the context of overt observation, as in the present scenario (Buck, Losow, Murphy, & Costanzo, 1992). Skin conductance decay data also confirmed huge individual variability in regulating the guilt response after the emotion manipulation. Judgments on the adaptive/maladaptive nature of guilt should therefore also consider individual differences in emotion regulatory ability, which may be associated with both normal and pathological variation in well-being (Ochsner & Gross, 2005).

Finally, both studies described here were characterised by a notable difficulty in eliciting strong pride that was distinct from the neutral condition: While self-report ratings were indicative of significant increases in pride, Study 1's discriminant function analysis failed to distinguish between the Neutral and Pride conditions, and Study 2 failed to detect significant differences in neural activation between the Neutral and Positive IAT conditions. By comparison, strong guilt was elicited in both studies. These findings may perhaps be attributed to the fact that the emotion paradigms were more specifically tailored toward the elicitation of guilt than that of pride.

They may also, however, reflect innate differences in our degree of response activation to positive and negative stimuli (for a review, see Cacioppo & Berntson, 1994). The "negativity bias" is a term coined to describe the phenomenon that organisms react more strongly (in terms of physiological, cognitive, behavioural, and emotional responses) to negative compared to positive or neutral stimuli, possibly because of evolutionary pressures to avoid harm (Cacioppo

& Gardner, 1999; Miller, 1961; Taylor, 1991). It is therefore also much easier to evoke negative than positive affect in experimental manipulations.

A different innate tendency, however, has been termed the "positivity offset", which refers to the stronger activation of (positive) approach versus avoidance motivational output at very low levels of arousal (Cacioppo & Gardner, 1999). The possible evolutionary benefit here may be that organisms are wired to approach novel objects in the absence of threat, which facilitates environmental exploration. In experimental situations, the positivity offset may manifest itself in biasing the evaluation of affectively neutral stimuli as mildly positive, thereby making it more difficult to discriminate between a positive and neutral condition. To overcome this inherent positivity offset, positive conditions should be designed to be sufficiently arousing, while neutral conditions may have to be slightly negative.

The Physiology of Guilt and Pride

Study 1 was a novel investigation of the physiology of a psychologically real guilt response. Results indicated that guilt is marked by strong physiological arousal and a cardiovascular response of reciprocal vagal withdrawal and cardiac sympathetic activation. The cardiac sympathetic effects (i.e., PEP shortening), however, continued longer and became more pronounced around 90 s after the emotion manipulation. By comparison, the pride response was characterized by mild cardiac reactivity and transient non-cardiac somatic arousal, as evidenced by the shift from HF to LF power. Because Pride participants displayed no PEP shortening (i.e., cardiac SNS arousal), the increase in LF power was interpreted as reflecting baroreflex-mediated vagal modulation (Moak et al., 2009), brought about by *somatic* SNS activation of vasomotor nerves.

The central finding of Study 1 was thus the distinct SNS activation patterns for guilt and pride. Although one would intuitively imagine that SNS-mediated arousal results in a unitary response, which is also the basic tenet of Walter Cannon's (1927) theory of undifferentiated autonomic arousal, mounting evidence points to the existence of discrete and highly patterned autonomic responses, also *within* the ANS axes (i.e., PNS and SNS) (Morrison, 2001). Accordingly, an extensive array of functionally distinct SNS responses may be distinguished to support diverse behavioural and physical challenges (Saper, 2002).

The present findings suggest that different organ functions may be modulated differentially during different emotional reactions, possibly to provide an adaptive advantage for the individual. Although previous research has suggested that different, organ-specific SNS activation patterns operate during mental compared to physical stressors (Wallin et al., 1992), the current results provide compelling evidence of differential SNS activity during discrete *emotional* reactions. Whereas the experience of guilt appeared to produce both cardiac and somatic SNS arousal, the current data suggested that SNS arousal associated with pride was not cardioselective in this way.

Few studies have looked specifically at changes in HRV, especially in the low frequency range, during discrete emotional states. The dearth of such HRV data may be attributed largely to the short measuring periods typically employed to capture an emotional response at its peak intensity, which is often inadequate for frequency analyses (Task Force, 1996). Nevertheless, my finding of respiratory-mediated vagal unloading during guilt and pride is consistent with previous reports detailing decreased RSA amplitude following exposure to either positive or negative emotional stimuli (Frazier et al., 2004; Rainville et al., 2006; Ravaja, 2004; Ritz, Alatupa, Thons, & Dahme, 2002). This response purportedly reflects competent engagement with the environment, enabling an individual to respond rapidly in either approach- or withdrawal-related fashion, without the need to deploy the slower-acting SNS (Porges, 1995, 2001). In a similar vein, Beauchaine (2001) argued that vagal withdrawal reflects nonspecific emotional responsiveness. Excessive vagal withdrawal, however, is nonadaptive and appears to be related to emotional lability of a fight-flight nature (Beauchaine, 2001; Yeragani et al., 1993).

The pride response was, arguably, physiologically less pressing than guilt: total HRV was not reduced in this condition, indicative of preserved physiological flexibility (Porges & Byrne, 1992). The pronounced shift to LF power in Pride participants also mimicked a response previously observed during the positive emotion of appreciation (McCraty, Atkinson, Tiller, Rein, & Watkins, 1995). In McCraty et al.'s appreciation condition, participants were required to sincerely feel appreciation or another positive emotion toward someone by focusing on their heart. A strong increase in spectral power was observed around 0.1Hz (Mayer waves), which is similar to the LF peak of 0.096 ± 0.02 Hz measured in the Pride participants here.

Pride's physiological response may also be argued to reflect its psychological function, namely to fuel and reinforce socially valued acts (Tracy & Robins, 2007a). Increases in relative

LF power at the expense of HF power in the cardiac spectrogram have been described recently as a measure of decreased chaos (i.e., less influence from HF breathing rate changes that are unpredictable or chaotic in nature) in the cardiovascular system (Wu et al., 2009). The lesser contributions of HF, and concomitant greater contributions of LF, to cardiac spectral power during pride should therefore decrease the homeostatic demands on the body and facilitate positive affect. In the present manipulation, pride thus served to decrease the inherently chaotic HF power but not the total power in HRV, and furthermore increased somatic SNS arousal. This transient non-cardiac SNS arousal in pride can be viewed as 'being in the zone', i.e., SNS-aroused but not stressed, relaxed yet focused – a pleasurable feeling that should encourage future pride-eliciting behaviours.

By comparison, the reduction in HRV across the TF band, coupled with the increased sympathetic activity experienced by the Guilt participants, resembles an overall stress response pattern (e.g., Friedman & Thayer, 1998). This response may have been augmented, however, by the experimental context. Stemmler (1989) has argued against absolute emotion specificity, promoting instead the view that the specific context counts. According to the component model of somatovisceral response organization (Stemmler, Aue, & Wacker, 2007; Stemmler et al., 2001), variation in an emotional response may be brought about by both the physical context as well as the contextual demands of the situation. The latter includes any motivational and behavioural demands necessitated by the momentary situation. The presence of another person, for example, may influence expressive behaviour because of learned display rules. In this regard, the presence of unfamiliar others has been shown to attenuate emotional expressivity, i.e., increased suppression (Buck et al., 1992).

Taking these considerations into account, it may be possible that participants in the Guilt condition experienced an amplified physiological response because they were unable to speak or to rectify the situation immediately. Suppression, or keeping affect (negative or positive) from being expressed, has often been reported to correspond with greater magnitudes of sympathetic arousal (Gross & Levenson, 1997; Roberts, Levenson, & Gross, 2008). This amplified response is thought to arise because curtailed emotional expression may be associated with effort and therefore greater energy expenditure (Ochsner & Gross, 2004). No amplified cardiovascular sympathetic arousal or physiological cost, however, was observed in the Pride participants, despite the fact that they also received the instruction to refrain from talking.

A final consideration in terms of the physiological responses associated with guilt concerns the fact that some emotions may be associated with a broader range of possible context-bound action sequelae than others, and therefore also with more variable patterns of autonomic activity (Lang et al., 1990). The complex emotion of guilt is likely to fall into this domain of emotions, such that the specific scenario may dictate the appropriate course of action to follow, and thus the concomitant ANS responses.

Physiological responses in Study 1 compared to those in Study 2. In contrast to Study 1, the Negative IAT/Guilt condition in the fMRI paradigm of Study 2 was associated with the lowest HR reactivity, while the Positive IAT/Pride condition was associated with the highest HR reactivity. As mentioned previously, the disparity between HR data from Study 1 and 2 may be explained by a careful consideration of differences between the two emotion elicitation paradigms.

The first important distinction between Study 1and 2 is that participants were in different body positions. Participants in Study 1 were in an upright position, which is associated with orthostatic (postural) stress and increased sympathetic tone, whereas participants in Study 2 were in a supine position, which is associated with more room in the vasculature (i.e., no orthostatic stress), higher vagal tone, and lower HR in general (Berntson, Cacioppo, & Quigley, 1993; Hatch, Klatt, Porges, Schroeder-Jasheway, & Supik, 1986). Because of these postural differences in the tonic autonomic control of the heart, it cannot be assumed that emotional responses in both studies would have had the same impact on measures of cardiac reactivity (Cacioppo et al., 1994).

A supine body position may furthermore affect behavioural motivation of an emotion, which is also associated with distinct physiological responses (Stemmler et al., 2007). Higher levels of motivational activation associated with preparation for action, for example, are accompanied by greater cardiac acceleration (Lang, Bradley, & Cuthbert, 1997). Harmon-Jones and Peterson (2009) recently showed that anger's characteristic greater left than right frontal cortical activity, which is associated with approach motivation (E. Harmon-Jones et al., 2009), was significantly reduced when participants were in a supine body position, compared to when they were sitting upright. The body position, however, did not affect participants' subjective experience: anger ratings did not differ between upright and supine body positions.

Taken together, an individual's body position may impact significantly on both physiological and neural response systems, and should be taken into account whenever emotional responses obtained during the assumption of different body postures are compared.

A second distinction between Study 1 and Study 2 involves differences in the physical surroundings of the two emotion manipulations. Compared to the quiet research laboratory of Study 1, the unfamiliarity of an MRI scanner, including the enclosed space and loud noises, may have had a marked influence on emotion physiological responses. Support for the fact that participants experienced these two environments differently stems from self-report ratings of anxiety obtained at baseline: Whereas average baseline anxiety ratings for Study 1 participants ranged from 2.88 to 3.50 (on a scale of 1 to 9) across experimental conditions, the average baseline anxiety rating for the fMRI study was 5.17 (also on a scale of 1 to 9). Higher baseline levels of arousal in the fMRI study may thus also have impacted on further autonomic arousal.

Finally, an important distinction between Study 1 and 2 involves the emotion manipulation itself. Study 1 was based on a 90-s interpersonal induction technique with high ecological validity. By comparison, the emotion manipulation in Study 2 depended on participants' perception of a 30-s visual feedback presentation, which may be more similar to paradigms where participants view emotive film footage/pictures. The emotion manipulation in Study 2, however, also had a strong *internal* component, because emotion elicitation depended on self-reflection, rather than explicit positive or negative visual stimuli. Nevertheless, Study 2 required visual orienting, which is associated with well-characterized autonomic correlates, including cardiac deceleration and moderate electrodermal increases (Graham, 1979; Lang et al., 1997), which would have influenced the overall autonomic response. Cardiac deceleration associated with orienting to an event or stimulus is regarded as an indication of heightened attention and sensory processing (Bradley, Codispoti, Cuthbert et al., 2001), and may reflect increased vagal efference to moderate sympathetic output in the service of enhanced attentional allocation (Porges, Doussard-Roosevelt, Portales, & Greenspan, 1996).

As mentioned previously, HR is closely associated with valence during the perception of emotional stimuli: Positive stimuli are associated with HR acceleration or modest deceleration, whereas negative stimuli are associated with greater HR deceleration (Lang et al., 1993). The main purpose of visual feedback in Study 2, however, was to provide participants with "performance-related" feedback. Performance monitoring has also been associated with immediate changes in HR. In this literature, negative feedback, which is associated with a feedback-related negativity (FRN) brain potential (Nieuwenhuis, Holroyd, Mol, & Coles, 2004), elicits greater HR slowing than positive feedback (Somsen, Van der Molen, Jennings, & van Beek, 2000). Van der Veen, Crone, and colleagues have shown that the cardiac response to performance-related feedback is determined by the affective valence of the feedback, especially when performance-based expectations are violated, and that HR slowing is associated with remedial action on later trials (Crone et al., 2003; van der Veen, van der Molen, Crone, & Jennings, 2004). This group has also demonstrated that the degree of cardiac slowing is highly context-sensitive, such that HR deceleration also reflects the information value of the feedback (Crone, Bunge, de Klerk, & van der Molen, 2005). Crone at el. (2005) suggested that the brain systems giving rise to error-related cortical activity may be the same neural substrates that also underlie heartbeat slowing, and may therefore center on rostral anterior cingulate structures (for a review, see van Veen & Carter, 2006).

It should be noted that the data presented above on feedback-related cardiac activity were based on *phasic* HR responses, and may therefore not apply directly to data from the current investigation. Fowles (1988) has also suggested, however, that HR is sensitive to the valence of the feedback signal (reward versus punishment), based on his observations of *tonic* HR levels.

In light of the above considerations, it is clear that the physiological activation in response to emotion manipulations in Study 1 and 2 were influenced by vastly different internal and external factors, which may account for the very different response patterns observed. Moreover, the data reviewed above strongly predict the decelerative HR trend observed during the negative IAT performance feedback. It is important to note, however, that subjective reports of guilt were not associated with this cardiac slowing during the Negative IAT condition; rather, there was a significant positive association between guilt and HR reactivity.

Lang and colleagues have described the HR response to aversive stimuli in terms of different stages that unfold sequentially (Bradley, Codispoti, Cuthbert et al., 2001; Lang et al., 1997). According to their model, the cardiac response is initially characterized by an attentional bradycardia, but evolves to acceleration once the stimulus is perceived as threatening or unpleasant, indicating heightened SNS arousal. Incorporating this thinking, it may be that the single HR measurement obtained in the current study also consisted of two stages. In this vein, participants who experienced more guilt may have manifested a higher second-stage HR acceleration (i.e., vagal withdrawal after the initial deceleration), than those who reported less guilt, given the fact that these response patterns may be affected by individual difference characteristics (Sánchez-Navarro, Martínez-Selva, & Román, 2006). Additional support for the notion that higher guilt was associated with higher cardiovascular arousal is gained from fMRI activation data, which is discussed in the next section. These assumptions are inconclusive, however, and will need further validation, especially in light of the limited physiological data obtained during the fMRI study, and the relatively small number of participants for whom good HR data were available.

Neural activation associated with physiological responses. The vast majority of studies employing functional neuroimaging techniques have had a strong focus on identifying neural correlates associated with higher-order human functions, such as decision-making, language, and memory. In the enthusiasm to better understand those cognitions that define us as human beings, however, the investigation of more low-level functions (e.g., those that control and represent bodily states), have received comparatively less attention (Critchley et al., 2003; Critchley, Wiens, Rotshtein, Ohman, & Dolan, 2004). Growing evidence from neuroimaging, lesion, and cortical stimulation studies provide compelling evidence for the notion that descending signals from higher cortical areas in humans are also intimately associated with cardiovascular control (Colivicchi, Bassi, Santini, & Caltagirone, 2004; Oppenheimer, Gelb, Girvin, & Hachinski, 1992; Williamson, McColl, & Mathews, 2003). Disentangling areas associated with afferent and efferent ANS responses from those associated with concurrent task- or emotion-related changes is a challenging undertaking, however.

In light of this relatively gray area in our understanding of cortical functioning, some neuroscience researchers have directed their recent efforts toward identifying brain regions involved in the modulation of PNS and SNS outflows to the body (e.g., Critchley, 2002; Critchley, Corfield, Chandler, Mathias, & Dolan, 2000; Critchley et al., 2003; Critchley, Melmed, Featherstone, Mathias, & Dolan, 2002; Lane et al., 2009; Saper, 2002; Williamson et al., 2006). These investigations highlight a network of forebrain regions, including the mPFC, ACC, insula, thalamus, and amygdala, with significant influence on autonomic arousal and cardiovascular control. The ACC, ventral mPFC, and insula, especially, have received considerable attention.

Whereas cognitive interpretations of rostral ACC function predominate in the brain-behaviour literature, recent lesion and imaging studies emphasize its role as modulator of peripheral autonomic arousal responses (Critchley, 2005; Critchley et al., 2003; Zahn, Grafman, & Tranel, 1999). ACC activation associated with autonomic arousal, however, does not consistently map onto a particular area of this large cortical structure. Functional imaging studies investigating cortical areas associated with ANS responses point to a functional dissociation between the subgenual and dorsal ACC (Critchley, 2004). Whereas the dorsal ACC typically activates during effortful or attention-demanding tasks, the subgenual ACC shows deactivation during such tasks (Paus, Koski, Caramanos, & Westbury, 1998; Raichle et al., 2001). For example, dorsal/supragenual ACC activation has been associated with increased HR, arterial blood pressure, and electrodermal activity during both mental and physical tasks, suggesting a role in modulating sympathetic outflow (Critchley et al., 2000; Critchley et al., 2003; Nagai, Critchley, Featherstone, Trimble, & Dolan, 2004). By this notion, prominent supraACC activity detected in the current Negative IAT/Guilt condition may also reflect increased SNS output (e.g., electrodermal activity) during the emotional response.

In contrast, the ventromedial PFC has been associated with vagal efferent outflow to the heart, such that suppression of activity in this area elevates cardiovascular arousal (Gianaros, Van Der Veen, & Jennings, 2004; Wong, Massé, Kimmerly, Menon, & Shoemaker, 2007). Ventromedial prefrontal areas (including ACC) furthermore have strong anatomical connections with subcortical autonomic control centers, permitting fast influence of the prefrontal cortex on the autonomic system (Barbas, Saha, Rempel-Clower, & Ghashghaei, 2003). The mediodorsal thalamic nucleus (activated in the present investigation) also forms part of this medial network relaying cortical visceromotor output to autonomic structures (Öngür & Price, 2000).

Although more ventral prefrontal areas have thus been implicated in PNS modulation, Matthews and colleagues found a slightly more dorsal area within the left perigenual ACC (-12, 37, 8) to be significantly correlated with high frequency HRV (Matthews, Paulus, Simmons, Nelesen, & Dimsdale, 2004). In their study, participants performed a counting Stroop task that was associated with task-induced changes in autonomic activity. Of particular interest to the present investigation is the fact that the area identified by Matthews et al. corresponds very closely to the pACC area also identified in the guilt condition of the current study (-7, 34, 15), suggesting that it too may have played a role in parasympathetic modulation. Moreover, Lane et al. (2009) recently identified a similar area within rostral ACC (-6, 48, 8) to be significantly correlated with HF HRV across all experimental conditions in their emotion elicitation study.

They suggested that this area may be involved in “coordinating autonomic adjustments associated with maintaining a self-focused mental state” during emotion or non-emotional tasks (Lane et al., 2009, p. 221). Given the significant negative linear relationship between pACC activity and subjective reports of guilt, it could be deduced that greater guilt was associated with increased withdrawal of parasympathetic tone.

As mentioned previously, parasympathetic withdrawal is readily associated with both positive and negative emotions. In addition, the association between guilt and vagal withdrawal is in line with data from Study 1, where subjective reports of guilt were significantly associated with various measures of HRV. The relation between decreased activity in the pregenual ACC and decreased vagal efference during guilt therefore appears to be plausible, yet this interpretation remains inconclusive in the absence of any explicit HRV and respiratory frequency data obtained during the fMRI paradigm.

An alternative functional account of pACC activity during guilt may involve the experience of affective pain. Sander and colleagues, for example, reported decreased perigenual ACC activity during a post-exercise ischemia condition, which was not associated with any significant HR changes (Sander, Macefield, & Henderson, 2010). Because decreased perigenual ACC activity could therefore not be associated with vagal withdrawal, they interpreted the fall in its activity as indicative of negative affect associated with muscle ischemia. This interpretation is considered in more detail later in the discussion.

Correlational data obtained in the current investigation indicated a significant positive association between neural activity in the pACC and right posterior insula, suggesting increased functional connectivity between these areas during the guilt induction. There is also considerable previous evidence indicating substantial anatomical connections between the ACC and insular cortices (Carmichael & Price, 1995; Taylor, Seminowicz, & Davis, 2009). Given the fact that both the pACC and right posterior insula were negatively correlated with self-reports of guilt, but positively associated with each other, it is possible that they were involved in a similar function during emotion evocation.

The insula is another area consistently implicated in the representation as well as modulation of cardiovascular autonomic states (Critchley, 2005; Williamson et al., 1997). It therefore not only receives and represents visceral afferent information from the body (Cechetto & Saper, 1987), but also plays an active role in cardiovascular regulation. Stimulation within discrete regions of the insula, for example, has been associated with distinct changes in autonomic responses, e.g., heart rate and blood pressure (Cechetto & Chen, 1990; Oppenheimer et al., 1992). The insula is thus considered part of a central command system, also involving the mPFC, which acts to interpret sensory input and elicit appropriate ANS adjustments (Williamson et al., 2006; Williamson et al., 2003).

Data from functional imaging studies, however, cannot distinguish between sensory representation of autonomic changes and visceromotor generation of such changes. Furthermore, insular involvement during changing autonomic states may reflect SNS arousal, PNS withdrawal, or a combination of the two. Zhang and Oppenheimer (1997), for example, detected both sympathoexcitatory and sympathoinhibitory neuronal units within the insula. Deciphering the precise function of enhanced insular activity in neuroimaging investigations is therefore an intricate task.

The right posterior insula is often associated with homeostatic and cardiovascular control (Hanamori, 2005; Zhang, Rashba, & Oppenheimer, 1998). Cardiac dysregulation and increased risk of cardiac arrhythmias, for example, have been reported after right insular stroke (Cheung, Hachinski, & Cechetto, 1997; Colivicchi et al., 2004). Oppenheimer et al. (1992) associated the right insula with central SNS regulation, because stimulation of the right insula in five epileptic patients was associated with tachycardia and hypertension. Several findings, however, point to a possible cardioinhibitory role of the right posterior insula, i.e., its association with *lower* HR reactivity. For example, Critchley et al. (2000) found a significant negative covariation between activity in the right posterior insula (42, -14, 4) and HR, such that lower activity in this area corresponded with raised HR. Similarly, Gianaros and colleagues (2004) found increased HR to be associated with decreased activity in the right posterior insula during various working memory tasks. Finally, a recent study directly demonstrated the cardioinhibitory effect of the right posterior insula in an epileptic patient with a potentially damaged right posterior insula (Al-Otaibi, Wong, Shoemaker, Parrent, & Mirsattari, 2010). Al-Otaibi et al.'s patient exhibited reduced HR responses as well as nearly absent right insular activation during a handgrip task, compared to control subjects. In particular, however, stimulation of his right posterior inferior (but not superior) insular cortex, during both rest and exercise, was accompanied by significantly suppressed HR responses. Notably, the particular area of stimulation in this study was more posterior to the area that was stimulated in the Oppenheimer et al. (1992) study, which may account for the different findings. Al-Otaibi et al. concluded that the posterior inferior sector of the right insula may serve an important cardioinhibitory role, i.e., parasympathetic modulation of HR.

Taken together, the data reviewed above suggest the right posterior insula has a strong association with cardiovascular regulation. Its architecture, however, appears to be complex, with distinct regional subdivisions subserving different ANS control functions (Al-Otaibi et al., 2010). Given the limited physiological data obtained during the present investigation, the precise functional significance of the right posterior inferior insula activation during guilt is difficult to pinpoint. Increased activity in this area (as well as in the pACC), however, coupled with the decreased HR reactivity during the guilt manipulation, appear most consistent with a cardioinhibitory effect. Partial support for this interpretation stems from the fact that both areas had a weak negative correlation with the overall HR reactivity during the Negative IAT condition (rs > -.32, ps < .26). These correlations also reached significance for HR data from Run 2 (rs > -.60, ps < .03).

Similar to the pACC, increased activity in the right posterior insula (i.e., increased cardio-regulatory control) was associated with reduced self-reports of guilt. Given that emotion regulation consists of both automatic and intentional processes that affect the emotional experience (Gross, 1998), HRV has been suggested to be an important resource that can be drawn upon to support emotion regulation (Thayer & Lane, 2009). Accordingly, increased vagal tone or parasympathetic inhibition during an emotional response may reflect enhanced autonomic flexibility, and therefore represent an individual's ability to modulate emotional responses to meet contextual demands (Appelhans & Luecken, 2006). By this notion, some participants in the guilt condition may have been better at inhibiting a less beneficial response in the scanner, for example, feeling very emotional and guilty, in favour of a more beneficial response. Not surprising, therefore, was the fact that increased activity in the right posterior insula was also associated with higher ratings of pride following the guilt induction.

Physiology of guilt and pride in terms of health. A growing body of research is delineating how psychological factors impact physical as well as mental health, with discrete emotions and their associated physiological responses considered key in this by now well-recognized link (Dickerson, Gruenewald et al., 2004; Kring & Bachorowski, 1999; Lerner et al., 2007; Steptoe & Brydon, 2009).

In terms of guilt's effects on health, the observed increase in SNS activity may be associated with health benefits (e.g., stimulation of the immune system) when experienced periodically and followed by rapid recovery (Mayne, 1999). However, emotion-associated SNS activation may also be detrimental to health when sustained or experienced chronically. Because the SNS is associated with energy mobilization, the energy demands on the system may become excessive with extended SNS activation, which in turn may lead to notable pathophysiological effects on the cardiovascular system (Fisher, Young, & Fadel, 2009). Compromised vagal function (decreased HRV), has also been associated with various psychopathologies and medical conditions when prolonged, because of reduced cardioinhibitory control (Thayer & Lane, 2007). Although parasympathetic withdrawal during situations of psychological or cognitive challenge is associated with behavioural flexibility, excessive vagal reactivity is associated with stress vulnerability (Porges, 1992), and may place those individuals with already low baseline RSA at particular risk for developing panic attacks (Beauchaine, 2001). Cacioppo (1994) emphasized, however, the importance of considering individual differences in cardiac reactivity, and highlighted the large interindividual variations in sympathetic and parasympathetic contributions to HR reactivity during psychological stress. His findings suggest that cardiovascular reactivity mediated by high cardiac sympathetic activation (PEP) is more strongly associated with altered immune function and disease pathogenesis than vagal cardiac reactivity (RSA). Individual differences in physiological responses may therefore be integral when considering possible associations between guilt and disease pathogenesis.

Nevertheless, a consideration of guilt's health consequences ultimately would entail a consideration of its duration and frequency: it is the physiological wear and tear that appears to be costly for the organism (Ryaff & Singer, 2003). Experiencing intense guilt chronically may therefore have cumulative effects on total "allostatic load", thereby increasing risk for cardiovascular disease as well as psychopathology (Harder, 1995; McEwen, 2008).

Guilt is typically viewed as a psychologically adaptive emotion (Tangney & Dearing, 2002). Although it is certainly not a pleasant emotion phenomenologically, its adaptability may stem from its prosocial virtues. Izard (1991, p. 355) argued that "guilt binds the person to the source of the guilt and does not subside without reconciliation that tends to restore social harmony." By this notion, the high immediate guilt-related anxiety observed in the present investigation may serve an adaptive function by stimulating impulses to make amends, which in turn, may speed cardiovascular recovery. The idea that corrective action is an important mechanism to relieve guilt-associated distress is evident in cases where the harm caused is irreparable or irreversible. For example, trauma guilt is often associated with heavily exacerbated and prolonged guilt (Kubany & Watson, 2003), and guilt has been shown to evoke self-punishment when there are no opportunities for remedial action (Nelissen & Zeelenberg, 2009).

Compared to guilt, shame is a more dejection-based emotional state and is associated with hiding or withdrawing to avoid further humiliation (Ferguson & Crowley, 1997). Several theorists have argued that guilt may become maladaptive and related to pathology when it becomes fused with shame (Tangney 2007), a sequence of events that may be instigated when one decides not to respond positively or prosocially to the guilt-inducing situation. This account is consistent with research indicating that beneficial guilt is possible only when people accept their failures and transgressions and take appropriate responsibility for their misdeeds. In contrast, maladaptive guilt is characterized by chronic self-blame and obsessive rumination over one's transgressions (Orth et al., 2006). In this regard, rumination has been proposed to mediate the effects of stress on health. Ruminative thoughts about past events, for example, have been associated with prolonged physiological reactivity and autonomic dysregulation long after the source of the stress (Ottaviani, Shapiro, Davydov, Goldstein, & Mills, 2009; Ottaviani, Shapiro, & Fitzgerald, 2011). Moreover, Ottaviani and colleagues' (2011) results showed that thinking about a negative event has cardiovascular consequences of a similar magnitude as the event itself, and is associated with increased negative affect and anxiety. It is thus clear why guilt associated with chronic rumination may be linked to pathogenic health consequences.

In terms of pride, my data provided support for a positive or adaptive physiological response: Whereas the increase in SCL and pronounced shift to LF power observed are indicative of somatic arousal and alertness, the moderate parasympathetic withdrawal is associated with optimal engagement and preparedness to respond (Beauchaine, 2001; Porges, 1995). Pride participants also exhibited virtually no increase in cardiac contractility, and faster recovery of vagal control than participants in the Guilt condition.

Positive emotions, in general, have received far less attention than negative emotions, yet empirical support for their health-promoting effects are accumulating (Tugade et al., 2004). Frederickson and Levenson (1998) have proposed that one purpose of positive affect may be to speed cardiovascular recovery from negative events. By this notion, positive emotion restores homeostasis by relieving an individual of the psychological and physiological sequelae of specific action tendencies associated with a foregoing negative emotion. According to Fredrickson's (2001) broaden-and-build theory of positive emotion, positive affect has the potential to reduce autonomic arousal, and, importantly, to broaden cognitive attention and stimulate flexible and creative behaviours.

Positive emotion is also protective, in that it is strongly associated with resilient adaptation and coping with stress and adversity (Curtis & Cicchetti, 2007). In this regard, both positive social interaction and positive individual traits, e.g., self-esteem, contribute to the protective nature of positive affect in resilient adaptation (Garmezy, 1985; Werner, 2000). Given pride's close association with genuine self-esteem, e.g., as a barometer of an individual's current social status and acceptance (Tracy & Robins, 2007a, 2007b), positive emotions of pride may be instrumental in positive life adaptation. One line of support for the positive health benefits of achievement success comes from an interesting study that assessed longevity of Academy Award-winning actors and actresses (Redelmeier & Singh, 2001). Results indicated a survival advantage of almost 4 years for actors that won an Academy Award versus actors that never got nominated. The study's findings could not be explained by other factors such as income, occupation, talent or chance; instead results pointed to the impact of higher status and success in relation to lower mortality rates.

The Neurobiology of Guilt

The most prominent brain activation observed during real-time guilt in Study 2 were that of the mesial prefrontal cortex (mPFC), including activation within the superior frontal gyrus and anterior cingulate cortex, predominantly on the left side. These were the only areas that showed significant activation at a corrected threshold in the whole-brain analysis, and therefore constitute a rather salient response. Determining the precise functional roles of these activations during a complex affective state like guilt, however, is no simple matter, given the plethora of studies and tasks that typically lead to mPFC activation. Few would disagree, however, that an overarching functional role of areas lining the medial wall of the frontal cortex involves social cognition (Amodio & Frith, 2006).

Putative roles of the mPFC in social cognition include its importance in the monitoring and attribution of mental states (Fletcher, Happe et al., 1995; Gallagher & Frith, 2003), empathic ability to take another's perspective (Lamm, Batson, & Decety, 2007), reasoning involved in complex moral decision-making (Greene et al., 2004), predicting the consequences of one's actions (Moll et al., 2007; Moll et al., 2008), conscious experience, initiation, or modulation of emotion (Reiman et al., 1997), inhibition of emotional experience (Ochsner, Ray et al., 2004), and the representation of information concerning oneself versus others (Kelley et al., 2002). The importance of the mPFC in social cognition is further supported by evidence from populations with impaired social functioning. For example, the PFC has been found to be dysfunctional in psychopaths or aggressive offenders (Blair, 2010; Raine et al., 1994), and is underactive in autistic individuals, who have impaired ToM (Castelli, Frith, Happe, & Frith, 2002).

Drawing on previous imaging data and meta-analyses of prefrontal and ACC function, I have tried to tease apart the functional significance of each of the mPFC activations detected in the whole-brain analysis. In the following sections, I elaborate on the putative roles of these medial prefrontal areas during guilt, and also discuss in more detail some areas detected in the mask analysis. In particular, the insular cortex is discussed as a prominent site for the instantiation of emotional feeling states.

Superior frontal gyrus activations. In terms of the more dorsal PFC activation, i.e., the superior frontal gyrus/supraACC (BA 8/9; -4, 28, 30), I have argued that this area, together with the ventrolateral PFC (BA 47), may be important in signaling inappropriate social behaviour and initiating alternative motor responses (Finger et al., 2006). Ventromedial PFC, in particular, appears to be sensitive to cues of punishment and negative emotional reactions from others, which in turn serves to alter current behaviour (Blair & Cipolotti, 2000; Kringelbach & Rolls, 2003, 2004). Dorsal/supraACC has also been associated with the experience of regret, a feeling of responsibility for harm to oneself because of personal decisions (Berndsen et al., 2004). Coricelli et al. (2005) found enhanced dorsal ACC activity (10, 24, 34) when participants discovered that an alternative action would have led to a more favourable outcome. They argued that enhanced ACC activity may serve to assess future outcomes, while evoking behavioural adjustments on subsequent trials.

The above account of the more dorsal, superior frontal gyrus activation during guilt is consistent with Amodio and Frith's (2006) proposed functional account of the posterior rostral medial frontal cortex. After considering data from a host of functional imaging investigations, Amodio and Frith concluded that this area may be important in behavioural regulation by monitoring the outcome of current actions, while continually updating representations of and evaluating the merit of future actions. A similar functional account of the PFC has also been proposed by Moll and colleagues (2007; see also Wood & Grafman, 2003), who argued that strong anterior prefrontal activation during moral emotions might reflect the appraisal of social outcomes and future actions. Prosocial emotions, in particular, may involve such appraisals, because of their association with socially-valued behaviours aimed at amending mistakes or social bonds. The dorsal superior frontal gyrus activation during guilt thus supports enhanced cognitive processing during social transgressions, which appears to be important for initiating alternative behaviours and evaluating possible future actions.

Given the substantial neurocognitive literature on the conflict-monitoring role of the related dorsal ACC region (e.g., Botvinick et al., 2004; Botvinick et al., 2001; Braver, Barch, Gray, Molfese, & Snyder, 2001), however, it is likely that the supraACC activity detected in the present investigation primarily serves to interrupt action (before initiating alternative motor responses). In this literature, conflict-related ACC activity is associated with increased reaction times and more controlled patterns of response on subsequent performance trials (van Veen et al., 2001), the successful inhibition of a prepotent response on No-Go trials in the Go/No-Go task (van Veen & Carter, 2002), and the slowing of performance following response errors (Gehring & Fencsik, 2001; Kerns et al., 2004). Moreover, cardiac deceleration in response to performance errors has been construed as the transient inhibition of ongoing action representations (Jennings & van der Molen, 2002). Conflict-monitoring via the ACC is therefore readily associated with the interruption of action (Amodio, Master et al., 2008). A region remarkably similar to the region identified in the present study (-4, 28, 30) is also consistently activated in studies of conflict-monitoring, compare, for example (-3, 32, 31) (van Veen et al., 2001), and (-2, 28, 31) (Botvinick, Nystrom, Fissell, Carter, & Cohen, 1999).

In light of the above, I therefore propose that activation in the superior frontal gyrus/supraACC during guilt first serves to stop current behaviour, before alternative motor responses are initiated to rectify the harm caused.

The second superior frontal gyrus activation (BA 9/32; -4, 43, 18) corresponds to an area within anterior rostral prefrontal cortex that has been described as having a “uniquely social cognitive function” (Amodio & Frith, 2006, p. 270). In particular, Amodio and Frith have suggested that this area mediates meta-cognitive processes that enable us to reflect on our feelings and intentions.

In recent years, a growing interest in neurocognitive investigations has involved the neural correlates associated with self-related or self-referential processing, which may be described as the degree to which individuals perceive concepts to be related to themselves. While the question of self has intrigued philosophers and psychologists throughout history (Damasio, 1999; Descartes, 1641/1993; James, 1892), functional neuroimaging is finally providing some testable hypotheses regarding the brain regions associated with self-related processing (Frith & Frith, 2003; Kircher et al., 2000; Northoff & Bermpohl, 2004).

In a meta-analysis of self-related neuroimaging studies, Northoff and colleagues (2006) highlighted the role of cortical midline structures (CMS) in self-referential processing. They described a network in which both ventral (including the pACC) and dorsal (supraACC and dorsomedial PFC) areas of the mPFC are key nodes in a neural system subserving explicit self-association. In terms of these anterior frontal areas, Northoff et al. (2006) expressed a similar view to that of Amodio and Frith (2006), namely that the commonly observed activation of anterior CMS structures during different emotions probably reflects the high degree of self-referentiality shared by all emotions, rather than inherent emotional processing occurring in his area. As a result, emotions that require less self-reflection may be associated with reduced mPFC activation (see, e.g., Britton, Phan et al., 2006; Takahashi et al., 2008). The involvement of the mPFC in self-reflection is also supported by evidence from neurological patients. Patients with frontotemporal lobar degeneration, a condition associated with neural loss in the mPFC, show profound impairments in the expression of self-conscious emotions, and therefore also processes related to self-appraisal (Sturm et al., 2006).

The posterior cingulate/precuneus is another key area identified by Northoff and colleagues (Northoff & Bermpohl, 2004; Northoff et al., 2006) as being involved in self-referential processing. The posterior cingulate responds when participants engage in self-reflection (Johnson et al., 2002), view personally familiar faces (Gobbini, Leibenluft, Santiago, & Haxby, 2004), and retrieve previous judgments related to the self (Lou et al., 2004). Because of dense connections with the hippocampus, Northoff et al. (2006) suggested that this region may be important for integrating present and past self-referential stimuli in a temporal context. In the present investigation, significant positive correlations between signal increases in the posterior cingulate, precuneus, and medial prefrontal cortex suggest that these areas were functionally connected during guilt. This finding provides compelling support for the interpretation that participants engaged in self-reflection during emotion-elicitation, given that increased functional connectivity between CMS has been associated with increased self-referential processing (Kjaer, Nowak, & Lou, 2002; Lou et al., 2004; Northoff et al., 2006).

Taken together, significant activation of (and correlation between) CMS in Study 2 may represent the high degree of self-association and explicit self-reflection that participants engaged in during the guilt condition. Among these midline structures, however, the anterior rostral frontal cortex may play a unique role. Amodio and Frith (2006) have suggested that this area may not be limited to reflections about our own subjective experience, but may also be involved in concerns about our reputation, or the image of ourselves in the minds of others. This perception of another's view of self has been referred to as reflected self-knowledge (Ochsner et al., 2005).

Anterior rostral frontal cortex may therefore support representations of how (we think) others think about us, which is of course an inherent aspect of all self-conscious emotions (Eisenberg, 2000). Although the self is the object of evaluation in self-conscious emotions, Leary (2004, 2007) stressed the view that these emotions arise from concerns about *others*' evaluation of the self, and therefore always involve a real or imagined audience. He argued that reflected appraisals, i.e., people's beliefs about how they are being evaluated by others, constitute a key feature of self-conscious emotions. I therefore propose that the second superior frontal gyrus activation (BA 9/32) during guilt reflects participants' engagement in self-conscious reflection, which involves reflecting upon themselves as well as the image of themselves in the minds of others.

The ACC and subjective evaluation of pain. The third activation in the whole-brain analysis corresponded to another area within the anterior cingulate cortex, namely the pregenual ACC (pACC). The cingulate cortex is a large structure that may be parsed into different cytoarchitectonic and functional subdivisions that subserve a vast array of specialized cognitions and bodily functions (Koski & Paus, 2000; Vogt et al., 1992). It is a structure considered to play a pivotal role in the integration of cognitive and affective information (Allman et al., 2001), the monitoring of subjective feeling states (Craig, 2002), and the generation of autonomic changes in the body, as discussed earlier (Critchley, 2005; Critchley et al., 2003).

Broadly speaking, the anterior cingulate is often segregated into two divisions (Bush et al., 2000): the dorsal division is principally involved in the mediation of cognitive processes (e.g., modulating attention and monitoring actions), whereas the rostral-ventral division, which lies around the genu of the corpus callosum, is associated with affective processing (e.g., inducing and regulating emotional responses) (Bush et al., 1998; Devinsky, Morrell, & Vogt, 1995; Mayberg et al., 1999; Whalen et al., 1998). Since the original dorsal/ventral model of ACC function has been proposed, however, a number of neuroimaging findings have been reported that cannot be accounted for by this model. In particular, neuroimaging studies of pain are not explained by the cognitive-affective account of ACC function (Eisenberger & Lieberman, 2004).

The experience of pain can be divided into two distinct dimensions that also map onto distinct brain regions, namely unpleasant sensory processing and emotional distress (Price, 2000). Neuroimaging studies of pain, however, point unexpectedly to the involvement of the dorsal/cognitive ACC division in the emotional distress component of physical, social, as well as empathic pain (Akitsuki & Decety, 2009; Coghill, McHaffie, & Yen, 2003; Eisenberger et al., 2003; Singer et al., 2004). To account for these observations, Eisenberger and Lieberman (2004) suggested that the function of the ACC may be better described in terms of a conflict monitoring system underlying both cognitive and emotional processing. By this model, the ACC has evolved as a neural substrate responsible for the detection of cues that may lead to both physical (e.g., pain, danger) and emotional (e.g., social separation) harm, which in turn signals the need for additional attention and coping mechanisms. In particular, Eisenberger and Lieberman proposed that dorsal ACC may be involved in sensorimotor representation of conflict (e.g., unanticipated pain), whereas rostral ACC may be more involved when the conflict has a specific object, person, or event at its focus (see also Satpute & Lieberman, 2006). Both rostral and dorsal ACC therefore subserve conflict detection, although they are sensitive to different forms of conflict.

The functional organization proposed by Eisenberger and Lieberman is conceptually consistent with Amodio and Frith's (2006) overarching model of medial frontal cortex function. Considering data from their meta-analysis, Amodio and Frith proposed that mental representations become more abstract and subjective as one moves from posterior to more anterior regions within the medial frontal cortex. This reasoning is well illustrated in the context of pain. Whereas objective, physical aspects of pain (e.g., temperature) is represented in more caudal regions of the ACC, the subjective experience of pain is represented in more anterior dorsal ACC regions (Davis, 2000; Rainville, Duncan, Price, Carrier, & Bushnell, 1997; Singer et al., 2004). Subjectively experiencing pain, however, is not the same as *thinking* about the pain. The most anterior rostral ACC appears to be involved in this meta-cognitive process. For example, in Singer et al.'s (2004) study, anterior rostral ACC (-6, 42, 18) was not active during the experience of personal pain, but was elicited only when *thinking* about someone else in pain. The most anterior regions of the mPFC/ACC therefore appear to be associated with more abstract representations of pain, for example, thinking about the subjective unpleasantness thereof.

Regarding the role of the pACC in the current guilt manipulation, I have argued that increased activation in this area may correspond to participants' down-regulation of their subjective experiences. Because the ACC is a neural substrate intimately associated with the experience and regulation of physical and social pain (Davis, 2000; Eisenberger et al., 2003), and because thinking about negative emotional experiences may be similar to thinking about pain (Amodio & Frith, 2006), an alternative interpretation of the role of the pACC during guilt evocation may involve the regulation of affective pain. In this regard, it is of interest that guilt shares significant overlap with the experience of pain on several fronts: Pain is also associated with affective unpleasantness, enhanced attentional resources, and the interruption of ongoing action (Wall, 1999). In addition, guilt and social pain are both associated with anxiety about possible social exclusion, threats to interpersonal relationships, and psychological distance from close others (Baumeister et al., 1995; Eisenberger & Lieberman, 2004; Izard, 1991; Panksepp, 1998). The link between guilt and pain thus appears to be plausible.

In light of the above, a different way to conceive of the negative association between pACC activity and subjective reports of guilt is to view increased ACC activity as subjectively 'blocking' one's pain. Support for this interpretation stems from the fact that increased rostral ACC has consistently been associated with a pain-relieving response mechanism during placebo and opioid analgesia (Petrovic, Kalso, Petersson, & Ingvar, 2002; Wager et al., 2004), stimulus-induced analgesia (García-Larrea et al., 1999), as well as hypnosis-induced changes in pain perception (Faymonville et al., 2000). For example, in the study by Petrovic et al. (2002), reduced self-reported pain during placebo analgesia was associated with increased activity in the pACC (18, 32, 14). In a similar vein, Salomons et al. (2004) proposed that rostral ACC plays a modulatory role in pain processing based on contextual information, such as the perceived controllability of the noxious stimulus, rather than processing information about the pain stimulus itself. Also consistent with this notion is rostral ACC's involvement in the cognitive modulation of pain: Increased activity in the perigenual ACC (-10, 32, -2) has been associated with reduced perception of painful stimuli during a cognitively demanding task (Bantick et al., 2002).

In summary, pACC activity in the current context may reflect a meta-cognitive regulatory response to affective pain. Accordingly, some participants responded to the guilt-induction by allowing themselves to experience pain, or being unable to cut themselves off from it (i.e., decreased pACC), while others were able to block out these feelings more effectively (i.e., increased pACC). As explained previously, such a regulatory response may have been mediated by increased vagal tone. This explanation is particularly appealing in light of Craig's (2003b) recent description of pain as a homeostatic emotion that also consists of a feeling and a motivation. By implication, individual differences in emotion regulatory capacity, as reflected in HRV, may thus also account for individual variation in pain sensitivity, or the ability to regulate pain (Appelhans & Luecken, 2006). Alternatively, individuals better able to regulate affective pain may perhaps have a more effective opioid system to relieve pain (Petrovic et al., 2002), given that placebo and opioid analgesia have a shared neural architecture, which is centered in the rostral ACC.

The insula and subjective feeling states. Several influential theories of emotion postulate a central role of peripheral autonomic feedback from the body in characterising the active emotional state, and by extension, motivational behaviour and self-awareness (Damasio, 1994, 1999; James, 1884; James, 1894; Lange, 1885). Various structures in the central nervous system have been implicated in the generation and representation of such changes in the internal bodily milieu (Saper, 2002). In an influential paper, Craig (2002) described the human insular cortex as the primary cortical site for the perception and integration of interoceptive (internal) information from the body, and its involvement in remapping these signals into conscious affective feelings. The insula has, in fact, often been referred to as the viscerosensory or limbic sensory cortex, because of its association with visceral sensation (Augustine, 1996; Penfield & Faulk, 1955).

As the primary interoceptive site, the posterior insula receives homeostatic afferent input from muscles and internal organs via the ventromedial thalamic nucleus, and is thought to represent this information in a topographic map of the body's internal state (Craig, 2002, 2003a). Sensory information of the physiological status of the body is then re-represented in the anterior insula, particularly on the right side, which is thought to subserve the subjective evaluation of the internal state (Craig, 2004; Critchley et al., 2004). Sensory information is also directly relayed to the ACC by way of the mediodorsal thalamic nucleus, to produce behavioural drive.

Whereas the anterior insula is therefore associated with the subjective awareness of the emotional feeling, the ACC is associated with autonomic and emotional control (i.e., limbic motor cortex). Because an emotion normally consists of both a motivation and concurrent feeling state (Rolls, 1999), the insula and ACC are believed to work together to create an integrated emotional response with concomitant autonomic effects (Craig, 2002). In a recent perspective, Craig (2009) further outlined the function of the anterior insula (often including the frontal operculum) as not only subserving the subjective evaluation of bodily states, but as a unique cortical substrate that instantiates *all* subjective feeling. By comparison, increased ACC activity associated with emotional responding may facilitate or interrupt ongoing behaviour, whereas the absence of such activity may contribute to decreased motivational performance (Critchley et al., 2003).

In light of the above, it is reasonable to assume that the left anterior insula and ventrolateral PFC activation observed in the current study reflects the subjective feeling state associated with guilt. Seeley et al. (2007) found activity in the left anterior insula to be associated with both an ‘executive control’ network and an emotional ‘salience’ network, which suggests that this area may underlie bidirectional awareness of emotion and cognitive functions. Montague and Lohrenz (2007) also recently provided an intriguing account of the function of the anterior insula that is consistent with my findings for guilt. They suggested that activity in the anterior insula reflects natural error signals in response to norm violations, which are experienced by the individual as unique negative feelings of disgust, pain, or guilt. These error signals serve to direct the individual to rectify the norm violations by adjusting subsequent behaviour.

Less clear, however, is the functional significance of the right posterior insula activation, and its negative linear relationship with self-reports of guilt. Given the literature on interoception reviewed above, it may be that significant activation in the right posterior insula during guilt reflects its involvement in the interoceptive representation of bodily afferent information. In contrast to the present finding, however, a more *dorsal* area of the mid/posterior insula is implicated as the primary projection area for visceral afferents (Craig, 2002; Craig, Chen, Bandy,

& Reiman, 2000). Moreover, an association with afferent bodily feedback would suggest a positive, rather than negative, association between activity in the right posterior insula and affective reports of guilt. The most plausible explanation of the right posterior insula's functional significance, therefore, is its putative involvement in cardiovascular control during the emotion manipulation, which was discussed earlier.

In terms of specific emotional feeling states, the left anterior insula activation observed in the current study is readily associated with aversive moral emotions, such as indignation or anger, and particularly, the intensity of disgust regardless of form (Harrison et al., 2010; Krolak-Salmon et al., 2003; Moll et al., 2007; Moll, de Oliveira-Souza et al., 2005; Zahn, Moll et al., 2009). Although disgust is predominantly associated with viscerosomatic reactions toward a class of natural stimuli, it is a complex emotional state also intimately associated with moral judgment and behaviour (Rozin, Haidt, & McCauley, 2000). In some situations, it therefore also has the characteristics of a 'moral emotion' (Izard, 1977; Miller, 1997).

The anterior insula activation during the current guilt condition may suggest a functional-anatomical correspondence between guilt and disgust or anger. This interpretation is supported by subjective self-report data from Studies 1 and 2, where participants commonly expressed both disgust and anger (directed toward the self) as emotions intimately associated with the experience of guilt. Results from Basile et al.'s (2011) study furthermore supported a strong connection between deontological guilt and activation in the left insula. The authors suggested that this form of guilt may be associated with strong aversive affect, such as moral disgust.

An important distinction between moral disgust and 'guilt-associated' disgust, however, is that the offending agent in interpersonal moral disgust is external, whereas the offending agent in guilt is internal or 'self', with disgust possibly directed toward one's bad behaviour. Disgust arising in interpersonal situations is associated with the restoration of purity, because the emotion initiates behaviours aimed at breaking off contact or removing the disgust-inducing entity (Haidt, 2003). By a similar notion, disgust associated with one's transgression of a social norm may serve to remove or renounce the bad behaviour in order to restore a baseline of purity.

Lack of amygdala activation. Because the current emotion manipulation was designed to be emotionally arousing, I anticipated observing significant activation in the amygdala, given its role in processing emotionally salient information (Liberzon et al., 2003; Phan, Taylor et al., 2003). The negative finding for amygdala activation, however, may be explained by a careful consideration of the literature. Before such consideration, however, it should be noted that this finding is unlikely to be a result of susceptibility artifact, because inspection of the cortex mask, which was created based on functional activation maps, indicated that the bilateral amygdala was included in the analysis.

The first plausible explanation involves the nature of the emotion stimulus, which has been shown to affect amygdala activation. The amygdala, for example, appears to respond differentially to internally- and externally-generated emotion. Whereas amygdala activity is readily detected in response to salient externally-cued perceptual (especially visual) stimuli (Phan et al., 2002; Zald, 2003), it may be less engaged during internally-generated emotional responses, for example, recalling a previous life episode (Damasio et al., 2000; Reiman et al., 1997). Consistent with this notion, amygdala activation was not detected in previous PET studies of recall-based guilt (Shin et al., 2000), sadness, happiness or disgust (George et al., 1996; Lane, Reiman et al., 1997; Mayberg et al., 1999), or anger (Dougherty et al., 1999). While participants in the current study were presented with visual emotion eliciting stimuli (i.e., IAT feedback), the emotional response depended on the mental processing of this information, the application of this information to the ‘self’, and possibly the recollection of previous personal episodes of prejudice. The paradigm therefore did not strictly conform to criteria for externally-generated emotion elicitation.

The second plausible task-related factor that may have influenced amygdala activity, is top-down cognitive control (Phan, Wager et al., 2004). Tasks requiring increased cognitive effort (e.g., reappraisal of stimulus content for self-relatedness) may attenuate amygdala/limbic activity (Ochsner, Bunge, Gross, & Gabrieli, 2002; Ochsner, Ray et al., 2004). In contrast, passive emotion conditions (i.e., simply viewing emotional stimuli without performing an additional task) are more readily associated with increased amygdala activity (e.g., Harenski & Hamann, 2006; Lange et al., 2003; Moll, de Oliveira-Souza, Eslinger et al., 2002). Given the significant cognitive component of the present paradigm, and evidence suggesting that some participants regulated their emotional responses, it is likely that these factors also influenced amygdala activity.

The third plausible explanation involves the fact that amygdala detection may be influenced by significant habituation effects. For example, the amygdala demonstrates rapid decreases in activation in response to repeated presentation of emotive stimuli (Liberzon et al., 2000). Furthermore, a response may occur transiently at the start of an emotion stimulus, rather than being sustained throughout stimulus presentation (Siegle, Steinhauer, Thase, Stenger, & Carter, 2002; Zald, 2003). In this regard, more time-consuming induction methods (e.g., emotional recall) may be more susceptible to within-experiment amygdala habituation effects than methods employing direct perceptual induction (Breiter et al., 1996). Such habituation effects may arise because the amygdala is considered to be more active during the emotion induction phase, than during the feeling-phase (Büchel, Morris, Dolan, & Friston, 1998). Given the temporal characteristics of the current paradigm (i.e., the use of a 30-s block design), sustained amygdala activity over an extended period of time would have been necessary to observe significant activation in this structure.

In summary, the literature reviewed above does not preclude the amygdala from being active during guilt, but strongly suggests that the detection of amygdala activity during the current guilt induction may have been hampered by several factors.

Behavioural Motivation of Guilt

Guilt has been described as a form of anxiety associated with the threat of social exclusion due to misconduct (Baumeister et al., 1995). This anxiety is thought to be adaptive in that it promotes reinforcement learning, while simultaneously inhibiting the transgressive behaviour (Monteith, 1993; Monteith et al., 2002). In this way, guilt functions as a punishment cue to help the individual respond more carefully in future (Devine et al., 1991). Making amends, however, has been suggested as a line of action that may relieve or resolve guilt-associated distress and "give rise to positive-affect-evoking cognitions that counteract the negative affect" (Izard, 1991; Kubany & Watson, 2003, p. 77).

Whereas guilt proneness has been associated with BAS sensitivity and approach orientation (Sheikh & Janoff-Bulman, 2010), the current findings across both Studies in this thesis suggest that high BIS-sensitive individuals, as measured by the Carver and White (1994) BIS/BAS scales, will experience stronger guilt when they find themselves in a real guilt-evoking scenario. This finding provides support for the idea that initial guilt functions as a punishment cue (Monteith, 1993), and suggests that current guilt is distinct from questionnaire constructs of guilt, which are typically employed to assess guilt-related phenomena.

Gray (1982) proposed that individuals high in BIS sensitivity will be sensitive to aversive events or punishment cues, and that this sensitivity will result in greater negative affect (in this case, guilt), as well as increased arousal and the momentary interruption of ongoing action, during specific situations of threat. As mentioned in the Literature Review, the association of the Carver and White BIS scale with conflict monitoring and the interruption of action (rather than behavioural avoidance), has been supported recently by findings from Amodio and colleagues (2008). I therefore agree with previous accounts, and argue that guilt-related anxiety, together with guilt cognitions, functions as a punishment cue that inhibits ongoing behaviour and encourages reparatory behaviour in order to alleviate distress (both personal as well as distress caused to another). This view is consistent with the notion that guilt facilitates a multifaceted self-regulatory process, starting with behaviour inhibition and transforming into approach-oriented, conciliatory behaviour when an opportunity for amendment appears (Amodio et al., 2007).

The specific nature of the self-regulatory process activated by guilt has been described in more detail in relation to prejudice reduction (Monteith, 1993; Monteith et al., 2002). According to this literature, heightened guilt, which is experienced as punishment, is fundamental in increasing an individual's motivation to reduce personal prejudice. In particular, Monteith (1993) presented and tested the following model: Low-prejudice people with well-internalised, nonprejudiced beliefs experience conflict when they transgress a personal standard. Awareness of the transgression is accompanied by the elicitation of discrepancy-associated affect, i.e., guilt, which briefly interrupts ongoing behaviour and motivates a series of responses to help the individual respond more effectively in future. The latter includes behaviours aimed at identifying environmental stimuli that triggered the discrepant response, e.g., increased attention and exploratory-investigative behaviours, and serves to establish cues for control (Monteith et al., 2002). Importantly, self-focus is significantly increased in the process.

Increased self-focused attention is also central to objective self-awareness theory, another theory on discrepancy-related cognitions and affects, though unspecific in terms of the specific nature of the ensuing negative affect (Duval & Silvia, 2002; Duval & Wicklund, 1972). According to this "self theory", increased self-awareness initiates greater comparison between the self and standards the self ought to have, which can motivate behaviours to change the self when discrepancies occur (e.g., Duval & Lalwani, 1999). Importantly, the increased self-focus is key to the motivation to be congruent with self standards, because it increases the accessibility as well as the significance of the discrepancy (Phillips & Silvia, 2005). Indirect evidence for a state of increased self-awareness following guilt-induction may be derived from the present fMRI study: During the Negative fixation period, increased activity in the mPFC and posterior cingulate (i.e., cortical midline structures) could be argued to reflect participants' increased engagement in self-reflection (Northoff et al., 2006). Participants therefore appeared to have heightened self-focus even after the emotion induction occurred.

Taken together, the events triggered by guilt constitute an inhibitory, self-regulatory system, which causes future responses to be slowed or inhibited through controlled processing (Monteith, 1993). This description shares significant overlap with Gray's (1981, 1982) initial neuropsychological model of the BIS, which states that discrepant responses are associated with the formation of response-contingent punishment cues that are crucial for the organism to gain control over (i.e., inhibit) future discrepant responses. The interruption of ongoing action, a hallmark of BIS activity (see e.g., Patterson, Kosson, & Newman, 1987), was in fact directly evident in Study 1, where one participant took an extended amount of time to carry on with unfinished tasks following the guilt induction. Because guilt's proposed self-regulatory mechanisms in the prejudice literature also have the effect of reducing prejudice and responding more appropriately or prosocially in future scenarios (Monteith, 1993; Monteith et al., 2002), this theorizing is consistent with guilt's proposed adaptive properties (cf. Tangney & Dearing, 2002).

Another salient finding from the present investigation was that self-reported guilt was negatively associated with BAS in Study 1 and in Study 2, which suggests that low BAS-sensitive individuals were more likely to report heightened guilt. Given the predictive value of BAS on both anger and self-reported frustration (Carver, 2004; Harmon-Jones, 2003), it may be that high BAS-sensitive individuals were more inclined to feel angered or frustrated, rather than guilty, by the guilt manipulations.

A different interpretation relates to previous research suggesting strong BAS individuals to be insensitive to BIS warnings or cues of punishment, focusing their attention instead on cues of incentive (Patterson et al., 1987; Patterson & Newman, 1993). In this regard, Corr (2001) has proposed the view that the BIS and BAS have both facilitatory and antagonistic effects: facilitatory in the sense that BIS mediates responses to aversive cues and BAS mediates responses to appetitive cues, and antagonistic in the sense that high BIS antagonizes or impairs BAS behaviours, and vice-versa. According to this joint subsystems account of motivational behaviour, state appetitive and aversive motivation is influenced by *both* the BIS and the BAS. Accordingly, state negative emotion, for example, should theoretically be highest in high BIS and low BAS individuals.

BAS scores in the current fMRI study were also positively associated with increased activity in the pACC and the right posterior insula. In light of the postulated functional roles of these areas described earlier, it is possible that high BAS-sensitive individuals were better at regulating negative affect, or more insensitive to subjective pain. In addition, these effects may have been mediated in part by decreased emotional state reactivity (i.e., reduced vagal unloading) to guilt-evoking stimuli. In this regard, much evidence associates antisocial impulsivity (i.e., a proposed overactive BAS) with physiological underarousal (see, e.g., Fowles, 2000; Knyazev et al., 2002). These are intriguing possibilities that will need further investigation.

Physiological parameters of BIS and BAS responding. Several theorists have argued convincingly for physiological markers of BIS and BAS activation during tasks that are designed to activate these motivational tendencies (Beauchaine, 2001; Fowles, 1980). Putative physiological correlates of BIS and BAS, however, do not always replicate well across studies and have been found to be largely uncorrelated with self-reported BIS and BAS (Brenner et al., 2005; Heponiemi et al., 2004; Knyazev et al., 2002; Morgan, 2006). Because of these considerations, the current physiological data in terms of BAS and BIS activation are interpreted with caution.

Before proceeding, I briefly consider relevant changes to the original BIS theory. The revised theory makes strong distinctions between anxiety, which is mediated by BIS, and fear, which is mediated by a separate fight/flight/freeze system (FFFS) (McNaughton & Corr, 2004). According to these current theoretical accounts, BIS is activated when an individual experiences conflict between competing motivational tendencies (e.g., approach-avoidance, approach-approach, or avoidance-avoidance), and is therefore often activated by the simultaneous activation of the FFFS and BAS (when they are of similar intensity). BIS activation has the effect of inhibiting prepotent behaviour, but also induces anxiety and increased autonomic arousal, which facilitates behaviours like risk assessment. BIS is thought to resolve the conflict by either engaging the BAS or FFFS, based on evaluation of the reinforcing signals received (Smillie et al., 2006).

In the case of guilt, the acute emotional reaction intuitively activates both the FFFS (i.e., freezing in response to threat detection) and BAS (i.e., behavioural activation or active avoidance); it therefore follows that conflict between these systems would result in increased BIS activation. This description is consistent with the results for guilt from Study 1: it was associated with increased punishment sensitivity (i.e., FFFS), increased BAS activity (i.e., PEP; Brenner et al., 2005), increased BIS activity (i.e., SCL; Fowles, 1980), and increased arousal (i.e., RSA; Frazier et al., 2004). In addition, anxiety was the second most highly-rated emotion in the Guilt condition. Hypothetically speaking, the competition between BAS-mediated appetitive motivation and the fight/flight/freeze system may eventually be tipped in favour of approach for guilt (i.e., guilt-related amending behaviour), whereas it might be tipped in favour of avoidance in the case of shame (see, e.g., Dickerson, Gruenewald et al., 2004). This account of guilt's motivational direction is in line with Amodio et al.'s (2007) findings showing that guilt is initially associated with reduced approach motivation, followed by increased approach motivation during engagement in prosocial activity.

Neural correlates of the BIS. According to Gray and McNaughton (2000), the BIS depends on a hierarchical system of neural structures involving both subcortical (including the amygdala and septo-hippocampal system) and cortical structures (including the PFC and ACC). The assignment of functional roles to specific areas associated with BIS activation, however, is at present still relatively non-specific, particularly with regard to larger PFC and ACC structures (McNaughton & Corr, 2004). The neurocognitive correlates of BIS therefore remains to be elucidated in humans, yet strong conceptual similarities between the conflict-monitoring accounts of the ACC and the BIS has led to the proposal that BIS detects response conflicts and interrupts action via an ACC substrate (Amodio, Master et al., 2008).

The BIS, for example, has been described as a comparator that scans the environment for mismatches between actual and expected stimuli; when a mismatch occurs, it recruits additional cognitive resources and takes control over behaviour (Gray, 1982; Monteith et al., 2002). Similarly, the ACC is considered essential in detecting events or internal states that signal the need for enhanced attention and cognitive control (Botvinick et al., 2004). The conflict-monitoring account of ACC function furthermore highlights its use of error or punishment signals to trigger strategic adjustments in cognitive control and adapt future behaviour (see e.g., Amodio, Devine et al., 2008; Kerns et al., 2004). In a similar fashion, BIS activation results in more inhibited or controlled responses in future situations. Lastly, individual difference analyses have indicated that conflict-related ACC activity is amplified for individuals high in negative emotionality (Luu, Collins, & Tucker, 2000), a personality trait that corresponds to higher dispositional BIS sensitivity (Smillie et al., 2006).

In light of the above, the prominent conflict-related dorsal/supragenual ACC activity detected in the current fMRI study provides further support for the association between acute guilt and behavioural inhibition. This ACC activity, which may be associated with conflicting response tendencies, possibly serves to engage more regulative dorsal prefrontal areas (Botvinick et al., 2004). Interestingly, a much more left-lateralized, dorsal PFC activation was observed during the Negative fixation compared to the Negative feedback period, suggestive of more cognitive control post-emotion elicitation. Finally, ventrolateral PFC's association with punishment (e.g., Kringelbach & Rolls, 2004) supports the notion that guilt is experienced as a punishment cue.

Sex Differences

Consistent with past research on emotion, a female-only sample was used in both studies of this thesis to reduce possible sex-related variation in emotional responding, and to maximize the likelihood of intense emotional experiences (Shields, 1991). The use of a homogenous female sample was thought to facilitate the detection of distinct autonomic and neural activation patterns associated with guilt and pride. Although guilt and pride have been shown to function in the same way for men and women (Monteith et al., 2002; Tracy & Robins, 2008), it is important to consider potential sex differences and how they may relate to the present findings. In this regard, sex differences pertain to various aspects of the emotional response, including the experience (self-report), expression (nonverbal behaviours), and physiology (cardiovascular, immune, neural, etc.) thereof (Kring & Gordon, 1998).

There is a longstanding Western-based stereotype that women are more emotional than men (Fischer & Manstead, 2000). A relatively robust finding, for example, is that women are more emotionally expressive than men (Gross & John, 1998; Kring & Gordon, 1998). Women also typically report more intense emotional experiences (Alexander & Wood, 2000), and often react more sensitively than men, that is, they have stronger physiological reactions, especially in response to negative/aversive emotional stimuli (Bradley, Codispoti, Sabatinelli, & Lang, 2001;

Gard & Kring, 2007; Nater, Abbruzzese, Krebs, & Ehlert, 2006). Sex differences in emotion, however, are considerably more complex than outlined above, and are subject to great variability in experimental methods and measures as well as social factors e.g., gender roles and socialization histories (Brody, 1997; Grossman & Wood, 1993).

Considering guilt and shame, extensive research by Tangney and colleagues has indicated that females across all age ranges report a greater propensity to guilt and shame than their male counterparts (Tangney & Dearing, 2002). This observation may be reflected in findings of epidemiological studies, which typically show women to be at a higher risk for developing mood disorders, including anxiety and depression, than men (Sachs-Ericsson & Ciarlo, 2000). Women are also more prone to ruminate about the causes and consequences of their own negative moods (Nolen-Hoeksema, Morrow, & Fredrickson, 1993). Others, however, contend that it is not sex per se, but a particular gendered self-concept that mediates the difference. Evans (1984), for example, found feminine gender-typed individuals to experience higher levels of guilt than their masculine gender-typed counterparts.

In a study assessing both guilt-proneness and dispositional guilt between male and female college students, Benetti-McQuiod and Bursik (2005) found females to be more guilt-prone than males. Males, on the other hand, reported more dispositional guilt. The authors argued that men may be more likely to find themselves in actual guilt-inducing situations, because men may be less sensitive to others' emotions and therefore less aware of the impact of their own behaviours. Females, however, are likely to be more attuned to the consequences of their behaviours and therefore better at anticipating other's reactions, which, in turn, may lead them to be more effective at avoiding punishing situations. The latter female personality trait may also be associated with higher BIS sensitivity, an interpretation consistent with Carver and White's (1994) original finding of higher BIS scores for females compared to males. Sex differences in guilt and shame, however, have been inconsistent across studies (see e.g., Ferguson & Crowley, 1997), an issue possibly inflated by the confounding effects of measures assessing different facets of guilt and shame (Benetti-McQuoid & Bursik, 2005; Eisenberg, 2000), as well as different characteristics of the samples assessed.

With regard to pride, evidence suggests that this is an emotion expressed more frequently or intensely by males (Collins & Frankenhaeuser, 1989). Pride has become associated with maleness because it is thought to convey status to one's social group, and males typically enjoy higher status in Western (and many non-Western) societies than women (Tracy & Robins, 2008). Conversely, women generally have lower self-esteem than men (Robins, Trzesniewski, Tracy, Gosling, & Potter, 2002), and may therefore experience pride less frequently and less intensely – an interpretation consistent with the present findings. It is impossible, however, to determine whether the lower emotional intensity associated with pride in the present research (especially in Study 2) was a result of the emotion manipulation method, or sex-related differences in emotional responding.

Sex differences in the physiological domain of emotion also remain unclear (Brody & Hall, 2000), with a relative dearth of studies exploring such differences. Women generally respond to laboratory stressors with greater HR reactivity (Brenner et al., 2005; Heponiemi et al., 2004), and show greater electrodermal activity to unpleasant or sad events (Kring & Gordon, 1998), while men typically show greater blood pressure responses (Steptoe, Fieldman, Evans, & Perry, 1996). Various studies have, however, failed to find these particular sex differences in physiological responses (Carrillo et al., 2001; Jones et al., 1996; Nater et al., 2006). In addition, sex differences in arousal have been shown to be specific to the task, as well as to the particular physiological measure (Brody, 1999). For example, one physiological measure (e.g., skin conductance magnitude) in a given scenario may show men to be more aroused, while another (e.g., facial electromyographic activity) may show women to be more aroused (Lang et al., 1993). Physiological differences are also affected by an individual's emotional style of expression (Cacioppo et al., 1992). Although it has often been reported that men are internalizers (i.e., heightened physiological arousal without overt emotional expression) and women are externalizers (i.e., overt emotional expression without corresponding physiological arousal) (Buck, Savin, Miller, & Caul, 1972), some evidence suggests that both men and women also fit within the generalizer framework (i.e., expressing emotions across physiological, facial, and verbal modalities) (Kring & Gordon, 1998). All of the above factors therefore prevent a simple extrapolation of physiological findings from the current female sample to males.

A possible confounding factor in the literature may be the emotion manipulation itself: researchers typically use the same type of emotion-inducing task for both men and women. Some evidence suggests, however, that sex differences in physiological arousal disappear when men and women are exposed to situations that they find equally important or arousing (Frodi, 1976). Men, for example, experience more shame and guilt when they fall short in situations associated with typical male gender roles, e.g., physical strength and work-related performance (Efthim, Kenny, & Mahalik, 2001). Accordingly, I do not expect men and women to have different physiological responses for guilt and pride per se, but there may be some differences in the emotion manipulations required to elicit these emotions to the same extent in both sexes. Nevertheless, a great deal more research is necessary to elucidate the psychophysiological mechanisms of guilt and pride in men and women.

Sex differences in emotional processing have also been demonstrated in functional neuroimaging studies of emotion (George et al., 1996; Hamann & Canli, 2004; Mak, Hu, Zhang, Xiao, & Lee, 2009b), although studies that directly compare male and female responses are not very common. Based on women's heightened responsiveness to emotional material in general, it has been postulated that women should also show stronger brain activation in emotional tasks than men (Wager, Phan, Liberzon, & Taylor, 2003). A meta-analysis conducted by Wager et al. (2003), failed to find any overall sex differences in activation, however. Instead, their findings suggested that sex effects for emotion at the neural level are complex and subject to regional specificity. In particular, men tended to show more activation in posterior sensory and left inferior frontal cortices, while women tended to show more activation in medial prefrontal and midline limbic structures, including the ACC and thalamus. While speculative, Wager and colleagues suggested that these differences in activation may reflect gender differences in emotional processing, such that males direct more attention to sensorimotor aspects of emotional stimuli to assess required action, whereas females focus more on the subjective feeling state engendered by the emotional stimuli.

A more detailed discussion of sex differences in emotional brain activation, however, is beyond the scope of this thesis. For the present purposes, it may suffice to assume that sex differences also exist at the neural level during guilt, although they are likely to be subtle.

Limitations and Directions for Future Research

Data from the present thesis support and expand on various theoretical aspects of guilt and pride's motivation and function, yet findings should be considered in light of some methodological limitations. First, the individuals who participated in both the psychophysiology and fMRI investigations consisted of a very homogenous sample of college students (i.e., White, female, young); caution is thus advised when extending these findings to other cultural/racial groups, as well as to older adults (see, e.g., Bierbrauer, 1992; Eid & Diener, 2001; Levenson, Carstensen, & Gottman, 1994).

In terms of the physiological profiles detected for guilt and pride, future studies will be necessary to replicate the current finding of differential SNS activation patterns, as well as to determine the possible influences that environmental factors, specific action programmes, and individual differences in affective style, may have in the autonomic patterning of these emotions. While the current data are consistent with the notion of distinct physiological arousal patterns for discrete emotions, Study 1 was not designed with an aim of confirming Jamesian emotion theory. To confirm autonomic specificity for the moral emotions, one would also have to demonstrate autonomic differentiation between similarly valenced moral emotions (Levenson, 2003a). A fruitful line of future investigation may therefore be to investigate autonomic differences between negative moral emotions (e.g., guilt, shame, and embarrassment) and between positive moral emotions (e.g., elevation, gratitude, and pride). Like others (e.g., Harmon-Jones et al., 2007; Herrald & Tomaka, 2002), I encourage researchers to investigate these emotions in vivo in situations of high ecological validity.

Another limitation, as discussed previously, was the difficulty in eliciting intense pride in research participants. Future studies may benefit from examining the pride response in relation to a moral act of goodwill, instead of the performance-based pride that is usually investigated in laboratory paradigms. This form of prosocial pride may have a more authentic emotional reaction as a result. Similarly, the fMRI paradigm was more closely associated with deontological than altruistic guilt. Guilt is thought to be triggered most powerfully in the context of communal relationships, especially if one’s harmful action also threatens this relationship (Baumeister et al., 1994). Developing an ecologically valid imaging paradigm to study interpersonal or altruistic guilt may therefore shed more light on affective disorders associated with abnormal guilt, e.g., depression (O'Connor et al., 2002). To improve the efficacy of such a paradigm, particularly for the fMRI context, researchers may benefit from developing paradigms or stimulus sets that are tailored for each individual participant (see, e.g., Mak et al., 2009a).

Finally, the finding that activity in the pACC was related to participants’ subjective reports of guilt opens up new fields of enquiry for future research. For example, given that increases in pACC activity may reflect individual differences in the efficiency of the cortical opioid system to modulate affective pain (Petrovic et al., 2002), and the fact that the social attachment system is also modulated by endogenous opioids (Nelson & Panksepp, 1998; Panksepp, 2005), it follows that dysregulation of this system in early childhood (e.g., as a result of separation-distress or trauma) may prevent effective guilt regulation in later life.

Summary and Conclusions

Guilt has been described as the emotion most essential in the development and regulation of moral behaviour (Izard, 1991). Although much has been written about guilt, relatively little is known about this moral emotion's biological substrates, especially because of various methodological challenges in its elicitation. The present thesis showed that guilt may be elicited reliably and studied in various experimental contexts through use of ecologically valid techniques. In particular, some findings from the present investigation clearly diverged from studies that employed only questionnaire measures to assess guilt. The importance of studying guilt in real-time is thus evident. In addition, the psychophysiology study (Study 1) demonstrated unequivocally that cognitive mind processes may be expressed in peripheral reactions also for the more complex moral emotions, and that investigation of such arousal patterns may contribute to our understanding of how moral emotions motivate behaviour and affect mental as well as physical health.

Study 1 also indicated that current experiences of guilt are associated with strong bodily arousal and cardiovascular responses mediated by concomitant vagal withdrawal and cardiac sympathetic activation. The neuroimaging investigation (Study 2) supported these findings, again showing significant activation in brain structures associated with heightened autonomic arousal, e.g., the ACC and insula. Both studies furthermore pointed to an association between acute guilt and behavioural inhibition, with high BIS (i.e., punishment-sensitive) individuals reporting more guilt. Guilt therefore appears to interrupt ongoing action and promote self-reflection, before becoming approach-oriented to repair the harm caused. This shift may, in fact, be a critical distinction that separates adaptive guilt from maladaptive guilt and shame.

Notably, both the psychophysiology and neuroimaging studies demonstrated that guilt is not a pleasant emotional experience, either cognitively or physiologically, and that it is associated with various other negative emotions, depending on the particular context. Its adaptive nature, therefore, has to be intimately associated with its prosocial virtues, which not only rectifies the harm caused to another, but also alleviates personal distress and possibly speeds cardiovascular recovery.

As a final point of consideration, it is important to stress that individual differences in the physiology and experience of guilt appear to be considerable, and that more research is necessary to define the characteristics under which guilt becomes maladaptive. These may shed light on disease pathogenesis in terms of cardiovascular as well as psychopathological disorders. Importantly, given the lack of validity often afforded by self-report measures of emotion, these relationships are perhaps best studied using ecologically valid paradigms of real-time guilt elicitation.

REFERENCES

Adolphs, R. (2003). Is the human amygdala specialized for processing social information? *Annals of the New York Academy of Sciences, 985*, 326-340.

Adolphs, R., Baron-Cohen, S., & Tranel, D. (2002). Impaired recognition of social emotions following amygdala damage. *Journal of Cognitive Neuroscience, 14*, 1264-1274.

Aguirre, G. K., Zarahn, E., & D'Esposito, M. (1998). The variability of human, BOLD hemodynamic responses. *NeuroImage, 8*, 360-369.

Akitsuki, Y., & Decety, J. (2009). Social context and perceived agency affects empathy for pain: An event-related fMRI investigation. *NeuroImage, 47*, 722-734.

Al-Otaibi, F., Wong, S. W., Shoemaker, J. K., Parrent, A. G., & Mirsattari, S. M. (2010). The cardioinhibitory responses of the right posterior insular cortex in an epileptic patient. *Stereotactic and Functional Neurosurgery, 88*, 390-397.

Alexander, M. G., & Wood, W. (2000). Women, men, and positive emotions: A social role interpretation. In A. H. Fischer (Ed.), *Gender and emotion: Social psychology perspectives* (pp. 189-210). New York: Cambridge University Press.

Allen, J. J., Chambers, A. S., & Towers, D. N. (2007). The many metrics of cardiac chronotropy: A pragmatic primer and a brief comparison of metrics. *Biological Psychology, 74*, 243-262.

Allison, T., Puce, A., & McCarthy, G. (2000). Social perception from visual cues: Role of the STS region. *Trends in Cognitive Sciences, 4*, 267-278.

Allman, J. M., Hakeem, A., Erwin, J. M., Nimchinsky, E., & Hof, P. (2001). The anterior cingulate cortex. The evolution of an interface between emotion and cognition. *Annals of the New York Academy of Sciences, 935*, 107-117.

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

Amaral, D. G. (2003). The amygdala, social behavior, and danger detection. *Annals of the New York Academy of Sciences, 1000*, 337-347.

Amodio, D. M., Devine, P. G., & Harmon-Jones, E. (2007). A dynamic model of guilt. Implications for motivation and self-regulation in the context of prejudice. *Psychological Science, 18*, 524-530.

Amodio, D. M., Devine, P. G., & Harmon-Jones, E. (2008). Individual differences in the regulation of intergroup bias: The role of conflict monitoring and neural signals for control. *Journal of Personality and Social Psychology, 94*, 60-74.

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

Amodio, D. M., Harmon-Jones, E., & Devine, P. G. (2003). Individual differences in the activation and control of affective race bias as assessed by startle eyeblink response and self-report. *Journal of Personality and Social Psychology, 84*, 738-753.

Amodio, D. M., Harmon-Jones, E., Devine, P. G., Curtin, J. J., Hartley, S. L., & Covert, A. E. (2004). Neural signals for the detection of unintentional race bias. *Psychological Science, 15*, 88-93.

Amodio, D. M., Kubota, J. T., Harmon-Jones, E., & Devine, P. G. (2006). Alternative mechanisms for regulating racial responses according to internal vs external cues. *Social Cognitive and Affective Neuroscience, 1*, 26-36.

Amodio, D. M., Master, S. L., Yee, C. M., & Taylor, S. E. (2008). Neurocognitive components of the behavioral inhibition and activation systems: Implications for theories of self-regulation. *Psychophysiology, 45*, 11-19.

Appelhans, B. M., & Luecken, L. J. (2006). Heart rate variability as an index of regulated emotional responding. *Review of General psychology, 10*, 229-240.

Arnett, P. A., & Newman, J. P. (2000). Gray's three-arousal model: An empirical investigation. *Personality and Individual Differences, 28*, 1171-1189.

Aron, A., Fisher, H., Mashek, D. J., Strong, G., Li, H., & Brown, L. L. (2005). Reward, motivation, and emotion systems associated with early-stage intense romantic love. *Journal of Neurophysiology, 94*, 327-337.

Aronson, E. (1968). The theory of cognitive dissonance: A current perspective. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 4, pp. 1-34). San Diego, CA: Academic Press.

Aronson, E., Ellsworth, P. C., Carlsmith, J. M., & Gonzales, M. H. (1990). *Methods of research in social psychology* (2nd ed.). New York: McGraw-Hill.

Arthurs, O. J., & Boniface, S. (2002). How well do we understand the neural origins of the fMRI BOLD signal? *Trends in Neurosciences, 25*, 27-31.

Augustine, J. R. (1996). Circuitry and functional aspects of the insular lobe in primates including humans. *Brain Research Reviews, 22*, 229-244.

Ausubel, D. P. (1955). Relationships between shame and guilt in the socializing process. *Psychological Review, 62*, 378-390.

Banaji, M. R., & Greenwald, A. G. (1995). Implicit gender stereotyping in judgments of fame. *Journal of Personality and Social Psychology, 68*, 181-198.

Bantick, S. J., Wise, R. G., Ploghaus, A., Clare, S., Smith, S. M., & Tracey, I. (2002). Imaging how attention modulates pain in humans using functional MRI. *Brain, 125*, 310-319.

Barbas, H., Saha, S., Rempel-Clower, N., & Ghashghaei, T. (2003). Serial pathways from primate prefrontal cortex to autonomic areas may influence emotional expression. *BMC Neuroscience, 4*, 25.

Barger, S. D. (2002). The Marlowe–Crowne affair: Short forms, psychometric structure, and social desirability. *Journal of Personality Assessment, 79*, 286-305.

Bargh, J. A., & Chartrand, T. L. (1999). The unbearable automaticity of being. *American Psychologist, 54*, 462-479.

Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind*. Cambridge: MIT Press.

Baron, A. S., & Banaji, M. R. (2006). The development of implicit attitudes. Evidence of race evaluations from ages 6 and 10 and adulthood. *Psychological Science, 17*, 53-58.

Bartels, A., & Zeki, S. (2004). The neural correlates of maternal and romantic love. *NeuroImage, 21*, 1155-1166.

Basile, B., Mancini, F., Macaluso, E., Caltagirone, C., Frackowiak, R. S., & Bozzali, M. (2011). Deontological and altruistic guilt: Evidence for distinct neurobiological substrates. *Human Brain Mapping*, 32, 229-39.

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology, 5*, 323-370.

Baumeister, R. F., Stillwell, A. M., & Heatherton, T., F. (1994). Guilt: An interpersonal approach. *Psychological Bulletin, 115*, 243-267.

Baumeister, R. F., Stillwell, A. M., & Heatherton, T., F. (1995). Personal narratives about guilt: Role in action control and interpersonal relationships. *Basic and Applied Social Psychology, 17*, 173-198.

Beauchaine, T. P. (2001). Vagal tone, development, and Gray's motivational theory: Toward an integrated model of autonomic nervous system functioning in psychopathology. *Developmental Psychopathology, 13*, 183-214.

Beauchaine, T. P., Gatzke-Kopp, L., & Mead, H. K. (2007). Polyvagal Theory and developmental psychopathology: Emotion dysregulation and conduct problems from preschool to adolescence. *Biological Psychology, 74*, 174-184.

Beauchaine, T. P., Katkin, E. S., Strassberg, Z., & Snarr, J. (2001). Disinhibitory psychopathology in male adolescents: Discriminating conduct disorder from attention-deficit/hyperactivity disorder through concurrent assessment of multiple autonomic states. *Journal of Abnormal Psychology, 110*, 610-624.

Bechara, A., Damasio, H., & Damasio, A. R. (2000). Emotion, decision making and the orbitofrontal cortex. *Cerebral Cortex, 10*, 295-307.

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Manual for the Beck Depression Inventory-II*. San Antonio, TX: Psychological Corporation

Beer, J. S., Heerey, E. A., Keltner, D., Scabini, D., & Knight, R. T. (2003). The regulatory function of self-conscious emotion: Insights from patients with orbitofrontal damage. *Journal of Personality and Social Psychology, 85*, 594-604.

Beer, J. S., & Keltner, D. (2004). What is unique about self-conscious emotions. *Psychological Inquiry, 15*, 126-129.

Benetti-McQuoid, J., & Bursik, K. (2005). Individual differences in experiences of and responses to guilt and shame: Examining the lenses of gender and gender role. *Sex Roles, 53*, 133-142.

Berndsen, M., van der Pligt, J., Doosje, B., & Manstead, A. S. R. (2004). Guilt and regret: The determining role of interpersonal and intrapersonal harm. *Cognition and Emotion, 18*, 55-70.

Berntson, G. G., Bigger, J. T., Jr., Eckberg, D. L., Grossman, P., Kaufmann, P. G., Malik, M., et al. (1997). Heart rate variability: Origins, methods, and interpretive caveats. *Psychophysiology, 34*, 623-648.

Berntson, G. G., Cacioppo, J. T., & Quigley, K. S. (1991). Autonomic determinism: The modes of autonomic control, the doctrine of autonomic space, and the laws of autonomic constraint. *Psychological Review, 98*, 459-487.

Berntson, G. G., Cacioppo, J. T., & Quigley, K. S. (1993). Cardiac psychophysiology and autonomic space in humans: Empirical perspectives and conceptual implications. *Psychological Bulletin, 114*, 296-322.

Berntson, G. G., Lozano, D. L., & Chen, Y. J. (2005). Filter properties of root mean square successive difference (RMSSD) for heart rate. *Psychophysiology, 42*, 246-252.

Berthoz, S., Armony, J. L., Blair, R. J. R., & Dolan, R. J. (2002). An fMRI study of intentional and unintentional (embarrassing) violations of social norms. *Brain, 125*, 1696-1708.

Berthoz, S., Grèzes, J., Armony, J. L., Passingham, R. E., & Dolan, R. J. (2006). Affective response to one's own moral violations. *NeuroImage, 31*, 945-950.

Bierbrauer, G. (1992). Reactions to violation of normative standards: A cross-cultural analysis of shame and guilt. *International Journal of Psychology, 27*, 181 - 193.

Blair, R. J. (1995). A cognitive developmental approach to mortality: Investigating the psychopath. *Cognition, 57*, 1-29.

Blair, R. J. (2010). Psychopathy, frustration, and reactive aggression: The role of ventromedial prefrontal cortex. *British Journal of Psychology, 101*, 383-399.

Blair, R. J., & Cipolotti, L. (2000). Impaired social response reversal. A case of 'acquired sociopathy'. *Brain, 123* 1122-1141.

Blair, R. J., Morris, J. S., Frith, C. D., Perrett, D. I., & Dolan, R. J. (1999). Dissociable neural responses to facial expressions of sadness and anger. *Brain, 122*, 883-893.

Boddaert, N., Chabane, N., Gervais, H., Good, C. D., Bourgeois, M., Plumet, M. H., et al. (2004). Superior temporal sulcus anatomical abnormalities in childhood autism: A voxel-based morphometry MRI study. *NeuroImage, 23*, 364-369.

Boiten, F. (1996). Autonomic response patterns during voluntary facial action. *Psychophysiology, 33*, 123-131.

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

Botvinick, M., Nystrom, L. E., Fissell, K., Carter, C. S., & Cohen, J. D. (1999). Conflict monitoring versus selection-for-action in anterior cingulate cortex. *Nature, 402*, 179-181.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review, 108*, 624-652.

Boyd, R., & Richerson, P. J. (2009). Culture and the evolution of human cooperation. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences, 364*, 3281-3288.

Bozeat, S., Gregory, C. A., Ralph, M. A., & Hodges, J. R. (2000). Which neuropsychiatric and behavioural features distinguish frontal and temporal variants of frontotemporal dementia from Alzheimer's disease? *Journal of Neurology, Neurosurgery, and Psychiatry, 69*, 178-186.

Bradley, M. M., Codispoti, M., Cuthbert, B. N., & Lang, P. J. (2001). Emotion and motivation I: Defensive and appetitive reactions in picture processing. *Emotion, 1*, 276-298.

Bradley, M. M., Codispoti, M., Sabatinelli, D., & Lang, P. J. (2001). Emotion and motivation II: sex differences in picture processing. *Emotion, 1*, 300-319.

Bradley, M. M., Cuthbert, B. N., & Lang, P. J. (1990). Startle reflex modification: Emotion or attention? *Psychophysiology, 27*, 513-522.

Bramel, D. (1968). Dissonance, expectation, and the self. In R. P. Abelson, E. Aronson, W. J. McGuire, T. M. Newcomb, M. J. Rosenberg & P. H. Tannenbaum (Eds.), *Theories of cognitive consistency: A sourcebook* (pp. 355-365). Chicago: Rand McNally.

Braver, T. S., Barch, D. M., Gray, J. R., Molfese, D. L., & Snyder, A. (2001). Anterior cingulate cortex and response conflict: Effects of frequency, inhibition and errors. *Cerebral Cortex, 11*, 825-836.

Breiter, H. C., Etcoff, N. L., Whalen, P. J., Kennedy, W. A., Rauch, S. L., Buckner, R. L., et al. (1996). Response and habituation of the human amygdala during visual processing of facial expression. *Neuron, 17*, 875-887.

Bremner, J. D., Narayan, M., Staib, L. H., Southwick, S. M., McGlashan, T., & Charney, D. S. (1999). Neural correlates of memories of childhood sexual abuse in women with and without posttraumatic stress disorder. *American Journal of Psychiatry, 156*, 1787-1795.

Brenner, S. L., Beauchaine, T. P., & Sylvers, P. D. (2005). A comparison of psychophysiological and self-report measures of BAS and BIS activation. *Psychophysiology, 42*, 108-115.

Brigham, J. C. (1993). College students' racial attitudes. *Journal of Applied and Social Psychology, 23*, 1933-1967.

Britton, J. C., Phan, K. L., Taylor, S. F., Welsh, R. C., Berridge, K. C., & Liberzon, I. (2006). Neural correlates of social and nonsocial emotions: An fMRI study. *NeuroImage, 31*, 397-409.

Britton, J. C., Taylor, S. F., Berridge, K. C., Mikels, J. A., & Liberzon, I. (2006). Differential subjective and psychophysiological responses to socially and nonsocially generated emotional stimuli. *Emotion, 6*, 150-155.

Brody, L. R. (1997). Gender and emotion: Beyond stereotypes. *Journal of Social Issues, 53*, 369-394.

Brody, L. R. (1999). *Gender, emotion and the family*. Cambridge, MA: Harvard University Press.

Brody, L. R., & Hall, J. A. (2000). Gender, emotion, and expression. In M. Lewis & J. M. Haviland-Jones (Eds.), *Handbook of emotions* (2nd ed., pp. 338-349). New York: Guilford Press.

Brosschot, J. F., & Thayer, J. F. (1998). Anger inhibition, cardiovascular recovery, and vagal function: A model of the link between hostility and cardiovascular disease. *Annals of Behavioral Medicine, 20*, 326-332.

Brown, J. D., & Marshall, M. A. (2001). Self-esteem and emotion: Some thoughts about feelings. *Personality and Social Psychology Bulletin, 27*, 575-584.

Brownley, K. A., Hurwitz, B. E., & Schneiderman, N. (2000). Cardiovascular psychophysiology. In J. T. Cacioppo, L. G. Tassinary & G. G. Berntson (Eds.), *Handbook of psychophysiology. Second Edition* (pp. 224-264). New York: Cambridge University Press.

Brychta, R. J., Shiavi, R., Robertson, D., Biaggioni, I., & Diedrich, A. (2007). A Simplified Two-Component Model of Blood Pressure Fluctuation. *American Journal of Physiology: Heart and Circulatory Physiology, 292*, H1193-H1203.

Büchel, C., Morris, J., Dolan, R. J., & Friston, K. J. (1998). Brain systems mediating aversive conditioning: An event-related fMRI study. *Neuron, 20*, 947-957.

Buck, R., Losow, J. I., Murphy, M. M., & Costanzo, P. (1992). Social facilitation and inhibition of emotional expression and communication. *Journal of Personality and Social Psychology, 63*, 962-968.

Buck, R. W., Savin, V. J., Miller, R. E., & Caul, W. F. (1972). Communication of affect through facial expressions in humans. *Journal of Personality and Social Psychology, 23*, 362-371.

Bush, G., Luu, P., & Posner, M. I. (2000). Cognitive and emotional influences in anterior cingulate cortex. *Trends in Cognitive Sciences, 4*, 215-222.

Bush, G., Whalen, P. J., Rosen, B. R., Jenike, M. A., McInerney, S. C., & Rauch, S. L. (1998). The counting Stroop: An interference task specialized for functional neuroimaging - validation study with functional MRI. *Human Brain Mapping, 6*, 270-282.

Butler, E. A., Wilhelm, F. H., & Gross, J. J. (2006). Respiratory sinus arrhythmia, emotion, and emotion regulation during social interaction. *Psychophysiology, 43*, 612-622.

Cacioppo, J. T. (1994). Social neuroscience: Autonomic, neuroendocrine, and immune responses to stress. *Psychophysiology, 31*, 113-128.

Cacioppo, J. T. (2004). Feelings and emotions: Roles for electrophysiological markers. *Biological Psychology, 67*, 235-243.

Cacioppo, J. T., & Berntson, G. G. (1994). Relationship between attitudes and evaluative space: A critical review, with emphasis on the separability of positive and negative substrates. *Psychological Bulletin, 115*, 401-423.

Cacioppo, J. T., Berntson, G. G., Larson, J. T., Poehlmann, K. M., & Ito, T. A. (2000). The psychophysiology of emotion. In M. Lewis & J. M. Haviland-Jones (Eds.), *Handbook of emotions* (2nd ed., pp. 173 - 191). New York: Guilford Press.

Cacioppo, J. T., & Gardner, W. L. (1999). Emotion. *Annual Review of Psychology, 50*, 191-214.

Cacioppo, J. T., Klein, D. J., Berntson, G. G., & Hatfield, E. (1993). The psychophysiology of emotion. In M. Lewis & J. M. Haviland (Eds.), *Handbook of emotions* (pp. 119 - 142). New York: Guilford Press.

Cacioppo, J. T., & Tassinary, L. G. (1990). *Principles of psychophysiology: Physical, social, and inferential elements*. New York: Cambridge University Press.

Cacioppo, J. T., Uchino, B. N., & Berntson, G. G. (1994). Individual differences in the autonomic origins of heart rate reactivity: The psychometrics of respiratory sinus arrhytmia and preejection period. *Psychophysiology, 31*, 412-419.

Cacioppo, J. T., Uchino, B. N., Crites, S. L., Snydersmith, M. A., Smith, G., Berntson, G. G., et al. (1992). Relationship between facial expressiveness and sympathetic activation in emotion: A critical review, with emphasis on modeling underlying mechanisms and individual differences. *Journal of Personality and Social Psychology, 62*, 110-128.

Canli, T., Zhao, Z., Desmond, J. E., Kang, E., Gross, J., & Gabrieli, J. D. (2001). An fMRI study of personality influences on brain reactivity to emotional stimuli. *Behavioral Neuroscience, 115*, 33-42.

Cannon, W. B. (1927). The James-Lange theory of emotion: A critical examination and an alternative theory. *American Journal of Psychology, 39*, 115-124.

Cannon, W. B. (1929). *Bodily changes in pain, hunger, fear and rage*. New York: Appleton.

Carmichael, S. T., & Price, J. L. (1995). Limbic connections of the orbital and medial prefrontal cortex in macaque monkeys. *Journal of Comparative Neurology, 363*, 615-641.

Carmichael, S. T., & Price, J. L. (1996). Connectional networks within the orbital and medial prefrontal cortex of macaque monkeys. *Journal of Comparative Neurology, 371*, 179-207.

Carrillo, E., Moya-Albiol, L., Gonzalez-Bono, E., Salvador, A., Ricarte, J., & Gomez-Amor, J. (2001). Gender differences in cardiovascular and electrodermal responses to public speaking task: The role of anxiety and mood states. *International Journal of Psychophysiology, 42*, 253-264.

Carver, C. S. (2004). Negative affects deriving from the behavioral approach system. *Emotion, 4*, 3-22.

Carver, C. S., & Scheier, M. F. (1990). Origins and functions of positive and negative affect: A control process review. *Psychological Review, 97*, 19-35.

Carver, C. S., & White, T. L. (1994). Behavioural inhibition, behavioural activation, and affective responses to impending reward and punishment: The BIS/BAS scales. *Journal of Personality and Social Psychology, 67*, 319-333.

Casebeer, W. D. (2003). Moral cognition and its neural constituents. *Nature Reviews Neuroscience, 4*, 841-846.

Casey, B. J., Thomas, K. M., Welsh, T. F., Badgaiyan, R. D., Eccard, C. H., Jennings, J. R., et al. (2000). Dissociation of response conflict, attentional selection, and expectancy with functional magnetic resonance imaging. *Proceedings of the National Academy of Sciences of the United States of America, 97*, 8728-8733.

Castelli, F., Frith, C., Happe, F., & Frith, U. (2002). Autism, Asperger syndrome and brain mechanisms for the attribution of mental states to animated shapes. *Brain, 125*, 1839-1849.

Cechetto, D. F., & Chen, S. J. (1990). Subcortical sites mediating sympathetic responses from insular cortex in rats. *The American Journal of Physiology, 258,* R245-255.

Cechetto, D. F., & Saper, C. B. (1987). Evidence for a viscerotopic sensory representation in the cortex and thalamus in the rat. *Journal of Comparative Neurology, 262*, 27-45.

Chee, M. W., Sriram, N., Soon, C. S., & Lee, K. M. (2000). Dorsolateral prefrontal cortex and the implicit association of concepts and attributes. *Neuroreport, 11*, 135-140.

Chen, Y., Martinez-Conde, S., Macknik, S. L., Bereshpolova, Y., Swadlow, H. A., & Alonso, J. M. (2008). Task difficulty modulates the activity of specific neuronal populations in primary visual cortex. *Nature Neuroscience, 11*, 974-982.

Cheung, R. T., Hachinski, V. C., & Cechetto, D. F. (1997). Cardiovascular response to stress after middle cerebral artery occlusion in rats. *Brain Research, 747*, 181-188.

Christie, I. C., & Friedman, B. H. (2004). Autonomic specificity of discrete emotion and dimensions of affective space: A multivariate approach. *International Journal of Psychophysiology, 51*, 143-153.

Chua, P., Krams, M., Toni, I., Passingham, R., & Dolan, R. (1999). A functional anatomy of anticipatory anxiety. *NeuroImage*, *9*, 563-571.

Coan, J. A., & Allen, J. J. B. (2003a). Frontal EEG asymmetry and the behavioral activation and inhibition systems. *Psychophysiology, 40*, 106-114.

Coan, J. A., & Allen, J. J. B. (2003b). Varieties of emotional experience during voluntary emotional facial expressions. *Annals of the New York Academy of Sciences, 1000*, 375-379.

Coan, J. A., & Allen, J. J. B. (Eds.). (2007). *The handbook of emotion elicitation and assessment*. Oxford: Oxford University Press.

Coghill, R. C., McHaffie, J. G., & Yen, Y. F. (2003). Neural correlates of interindividual differences in the subjective experience of pain. *Proceedings of the National Academy of Sciences of the United States of America, 100*, 8538-8542.

Colder, C. R., & O'Connor, R. M. (2004). Gray's reinforcement sensitivity model and child psychopathology: Laboratory and questionnaire assessment of the BAS and BIS. *Journal of Abnormal Child Psychology, 32*, 435-451.

Colivicchi, F., Bassi, A., Santini, M., & Caltagirone, C. (2004). Cardiac autonomic derangement and arrhythmias in right-sided stroke with insular involvement. *Stroke, 35*, 2094-2098.

Collet, C., Vernet-Maury, E., Delhomme, G., & Dittmar, A. (1997). Autonomic nervous system response patterns specificity to basic emotions. *Journal of the Autonomic Nervous System, 62*, 45-57.

Collins, A., & Frankenhaeuser, M. (1989). Stress responses in male and female engineering students. *Journal of Human Stress, 4*.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation. Design and analysis issues for field settings*. Chicago: Rand McNally.

Cools, R., Clark, L., Owen, A. M., & Robbins, T. W. (2002). Defining the neural mechanisms of probabilistic reversal learning using event-related functional magnetic resonance imaging. *The Journal of Neuroscience, 22*, 4563-4567.

Coricelli, G., Critchley, H. D., Joffily, M., O'Doherty, J. P., Sirigu, A., & Dolan, R. J. (2005). Regret and its avoidance: A neuroimaging study of choice behavior. *Nature Neuroscience, 8*, 1255-1262.

Corr, P. J. (2001). Testing problems in J. A. Gray's personality theory: A commentary on Matthews and Gilliland (1999). *Personality and Individual Differences, 30*, 333-352.

Cosmides, L., & Tooby, J. (2000). Evolutionary psychology and the emotions. In M. Lewis & J. M. Haviland-Jones (Eds.), *Handbook of emotions* (2nd ed., pp. 91-115). New York: Guilford Press.

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

Craig, A. D. (2003a). Interoception: The sense of the physiological condition of the body. *Current Opinion in Neurobiology, 13*, 500-505.

Craig, A. D. (2003b). A new view of pain as a homeostatic emotion. *Trends in Neurosciences, 26*, 303-307.

Craig, A. D. (2004). Human feelings: Why are some more aware than others? *Trends in Cognitive Sciences, 8*, 239-241.

Craig, A. D. (2009). How do you feel - now? The anterior insula and human awareness. *Nature Reviews Neuroscience, 10*, 59-70.

Craig, A. D., Chen, K., Bandy, D., & Reiman, E. M. (2000). Thermosensory activation of insular cortex. *Nature Neuroscience, 3*, 184-190.

Crawford, J. R., & Henry, J. D. (2004). The positive and negative affect schedule (PANAS): Construct validity, measurement properties and normative data in a large non-clinical sample. *British Journal of Clinical Psychology, 43*, 245-265.

Critchley, H. D. (2002). Electrodermal responses: What happens in the brain. *Neuroscientist, 8*, 132-142.

Critchley, H. D. (2004). The human cortex responds to an interoceptive challenge. *Proceedings of the National Academy of Sciences of the United States of America, 101*, 6333-6334.

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

Critchley, H. D., Corfield, D. R., Chandler, M. P., Mathias, C. J., & Dolan, R. J. (2000). Cerebral correlates of autonomic cardiovascular arousal: A functional neuroimaging investigation in humans. *The Journal of Physiology*, *523*, 259-270.

Critchley, H. D., Mathias, C. J., & Dolan, R. J. (2002). Fear conditioning in humans: The influence of awareness and autonomic arousal on functional neuroanatomy. *Neuron, 33*, 653-663.

Critchley, H. D., Mathias, C. J., Josephs, O., O'Doherty, J., Zanini, S., Dewar, B. K., et al. (2003). Human cingulate cortex and autonomic control: Converging neuroimaging and clinical evidence. *Brain, 126*, 2139-2152.

Critchley, H. D., Melmed, R. N., Featherstone, E., Mathias, C. J., & Dolan, R. J. (2002). Volitional control of autonomic arousal: A functional magnetic resonance study. *NeuroImage, 16*, 909-919.

Critchley, H. D., Wiens, S., Rotshtein, P., Ohman, A., & Dolan, R. J. (2004). Neural systems supporting interoceptive awareness. *Nature Neuroscience, 7*, 189-195.

Crone, E. A., Bunge, S. A., de Klerk, P., & van der Molen, M. W. (2005). Cardiac concomitants of performance monitoring: Context dependence and individual differences. *Cognitive Brain Research*, *23*, 93-106.

Crone, E. A., van der Veen, F. M., van der Molen, M. W., Somsen, R. J., van Beek, B., & Jennings, J. R. (2003). Cardiac concomitants of feedback processing. *Biological Psychology, 64*, 143-156.

Crosby, F., Bromley, S., & Saxe, L. (1980). Recent unobtrusive studies of Black and White discrimination and prejudice: A literature review. *Psychological Bulletin, 87,* 546-563.

Crowell, S. E., Beauchaine, T. P., Gatzke-Kopp, L., Sylvers, P., Mead, H., & Chipman-Chacon, J. (2006). Autonomic correlates of attention-deficit/hyperactivity disorder and oppositional defiant disorder in preschool children. *Journal of Abnormal Psychology, 115*, 174-178.

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology, 24*, 349-354.

Curtin, J. J., Lozano, D. L., & Allen, J. J. B. (2007). The psychophysiologal laboratory. In J. A. Coan & J. J. B. Allen (Eds.), *Handbook of Emotion Elicitation and Assessment*. Oxford: Oxford University Press.

Curtis, W. J., & Cicchetti, D. (2007). Emotion and resilience: A multilevel investigation of hemispheric electroencephalogram asymmetry and emotion regulation in maltreated and nonmaltreated children. *Development and Psychopathology*, *19*, 811-840.

Damasio, A. R. (1994). *Descartes' error: Emotion, reason, and the human brain*. New York: G. P. Putnam's Sons.

Damasio, A. R. (1999). *The feeling of what happens: Body and emotion in the making of consciousness*. New York: Harcourt Brace.

Damasio, A. R., Grabowski, T. J., Bechara, A., Damasio, H., Ponto, L. L. B., Parvizi, J., et al. (2000). Subcortical and cortical brain activity during the feeling of self-generated emotions. *Nature Neuroscience, 3,* 1049-1056.

Darwin, C. (1872). *The expression of the emotions in man and animals*. London: John Murray.

Davidson, R. J. (1994). Complexities in the search for emotion-specific physiology. In P. Ekman & R. J. Davidson (Eds.), *The nature of emotion: Fundamental questions* (pp. 237-242). New York: Oxford University Press.

Davidson, R. J., Jackson, D. C., & Kalin, N. H. (2000). Emotion, plasticity, context, and regulation: Perspectives from affective neuroscience. *Psychological Bulletin, 126*, 890-909.

Davidson, R. J., Pizzagalli, D., Nitschke, J. B., & Kalin, N. H. (2003). Parsing the subcomponents of emotion and disorders of emotion: Perspectives from affective neuroscience In R. J. Davidson, G. H & K. Scherer (Eds.), *Handbook of affective sciences* (pp. 8-24). New York: Oxford University Press.

Davies, R. R., Hodges, J. R., Kril, J. J., Patterson, K., Halliday, G. M., & Xuereb, J. H. (2005). The pathological basis of semantic dementia. *Brain, 128,* 1984-1995.

Davis, K. D. (2000). The neural circuitry of pain as explored with functional MRI. *Neurological Research, 22*, 313-317.

Davis, M., & Whalen, P. J. (2001). The amygdala: Vigilance and emotion. *Molecular Psychiatry, 6*, 13-34.

Dawson, M. E., Schell, A. M., & Filion, D. L. (2000). The electrodermal system. In J. T. Cacioppo, L. G. Tassinary & G. G. Berntson (Eds.), *Handbook of psychophysiology* (2nd ed., pp. 200-223). New York: Cambridge University Press.

de Geus, E. J. C., & van Doornen, L. J. P. (1996). Ambulatory assessment of parasympathetic/sympathetic balance by impedance cardiography. In J. Fahrenberg & M. Myrtek (Eds.), *Ambulatory assessment. Computer assisted psychological and psychophysiological methods in monitoring and field studies* (pp. 141-164). Berlin: Hogrefe & Huber.

de Quervain, D. J., Fischbacher, U., Treyer, V., Schellhammer, M., Schnyder, U., Buck, A., et al. (2004). The neural basis of altruistic punishment. *Science, 305*, 1254-1258.

de Vignemont, F., & Singer, T. (2006). The empathic brain: How, when and why? *Trends in*

Cognitive Sciences, 10, 435-441.

De Waal, F. B. M. (2001). *Tree of origin: What primate behavior can tell us about human social evolution* Cambridge: Harvard University Press.

De Waal, F. B. M. (2006). *Primates and philosophers: How morality evolved*. New Jersey: Princeton University Press.

Dearing, R. L., Stuewig, J., & Tangney, J. P. (2005). On the importance of distinguishing shame from guilt: Relations to problematic alcohol and drug use. *Addictive Behaviors, 30*, 1392-1404.

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behaviour*. New York: Plenum Press.

Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry, 11*, 227-268.

Dehaene, S., Posner, M. I., & Tucker, D. M. (1994). Localization of a neural system for error detection and compensation. *Psychological Science 5*, 303-305.

Descartes, R. (1993). *Meditations on first philosophy: In which the existence of God and the distinction of the soul from the body are demonstrated* (D. A. Cress, Trans. 3rd ed.). Indianapolis: Hackett Publishing Company, Inc (Original work published 1641).

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

Devine, P. G., & Monteith, M. J. (1993). The role of discrepancy-associated affect in prejudice reduction. In D. M. Mackie & D. L. Hamilton (Eds.), *Affect,cognition, and stereotying: Interactive processesin intergroup perception* (pp. 317-344). San Diego, CA: Academic Press.

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

Devine, P. G., Plant, E. A., Amodio, D. M., Harmon-Jones, E., & Vance, S. L. (2002). The regulation of explicit and implicit race bias: The role of motivations to respond without prejudice. *Journal of Personality and Social Psychology, 82*, 835-848.

Devinsky, O., Morrell, M. J., & Vogt, B. A. (1995). Contributions of anterior cingulate cortex to behaviour. *Brain, 118 (Pt 1)*, 279-306.

Dickerson, S. S., Gruenewald, T. L., & Kemeny, M. E. (2004). When the social self is threatened: Shame, physiology, and health. *Journal of Personality 72*, 1191-1216.

Dickerson, S. S., Kemeny, M. E., Aziz, N., Kim, K. H., & Fahey, J. L. (2004). Immunological effects of induced shame and guilt. *Psychosomatic Medicine, 66*, 124-131.

Dougherty, D. D., Shin, L. M., Alpert, N. M., Pitman, R. K., Orr, S. P., Lasko, M., et al. (1999). Anger in healthy men: A PET study using script-driven imagery. *Biological Psychiatry, 46*, 466-472.

Dovidio, J. F., Kawakami, K., & Gaertner, S. L. (2002). Implicit and explicit prejudice and interracial interaction. *Journal of Personality and Social Psychology, 82*, 62-68.

Durston, S., Davidson, M. C., Thomas, K. M., Worden, M. S., Tottenham, N., Martinez, A., et al. (2003). Parametric manipulation of conflict and response competition using rapid mixed-trial event-related fMRI. *NeuroImage, 20*, 2135-2141.

Dutton, D. G., & Lake, R. A. (1973). Threat of own prejudice and reverse discrimination in interracial situations. *Journal of Personality and Social Psychology, 28*, 94-100.

Duval, T. S., & Lalwani, N. (1999). Objective self-awareness and causal attributions for self-standard discrepancies: Changing self or changing standards of correctness. *Personality and Social Psychology Bulletin, 25*, 1220-1229.

Duval, T. S., & Silvia, P. J. (2002). Self-awareness, probability of improvement, and the self-serving bias. *Journal of Personality and Social Psychology, 82*, 49-61.

Duval, T. S., & Wicklund, R. A. (1972). *A theory of objective self-awareness*. New York: Academic Press.

Duvernoy, H. M. (1999). *The human brain: Surface, blood supply, and three-dimensional sectional anatomy* (Second, completely revised and enlarged ed.). New York: Springer.

Eckberg, D. L. (1997). Sympathovagal balance: A critical appraisal. *Circulation, 96*, 3224-3232.

Efthim, D. W., Kenny, M. E., & Mahalik, J. R. (2001). Gender role stress in relation to shame, guilt, and externalization. *Journal of Counseling and Development, 79*.

Eid, M., & Diener, E. (2001). Norms for experiencing emotions in different cultures: Inter- and intranational differences. *Journal of Personality and Social Psychology, 81*, 869-885.

Einstein, D., & Lanning, K. (1998). Shame, guilt, ego development, and the Five-Factor model of personality. *Journal of Personality, 66*, 555-582.

Eisenberg, N. (2000). Emotion, regulation and moral development. *Annual Review of Psychology, 51*, 665-697.

Eisenberger, N. I., & Lieberman, M. D. (2004). Why rejection hurts: A common neural alarm system for physical and social pain. *Trends in Cognitive Sciences, 8*, 294-300.

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

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion, 6*, 169-200.

Ekman, P. (1993). Facial expression and emotion. *American Psychologist, 48,* 384-392.

Ekman, P. (1994). Moods, emotions, and traits. In P. Ekman & R. J. Davidson (Eds.), *The nature of emotion: Fundamental questions* (pp. 56-59). New York Oxford University Press.

Ekman, P. (2003). *Emotions revealed*. New York: Times Books.

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology, 17,* 124-129.

Ekman, P., & Friesen, W. V. (1976). *Pictures of facial affect*. Palo Alto: Consulting Psychologists.

Ekman, P., Levenson, R. W., & Friesen, W. V. (1983). Autonomic nervous system activity distinguishes among emotions. *Science, 221*, 1208-1210.

Elliot, A. J., & Devine, P. G. (1994). On the motivational nature of cognitive dissonance: Dissonance as psychological discomfort. *Journal of Personality and Social Psychology, 67*, 382-394.

Elliott, R., Dolan, R. J., & Frith, C. D. (2000). Dissociable functions in the medial and lateral orbitofrontal cortex: Evidence from human neuroimaging studies. *Cerebral Cortex, 10,* 308-317.

Ellsworth, P. C. (1994). William James and emotion: Is a century of fame worth a century of misunderstanding? *Psychological Review, 101,* 222-229.

Ellsworth, P. C., & Tong, E. M. (2006). What does it mean to be angry at yourself? Categories, appraisals, and the problem of language. *Emotion, 6,* 572-586.

Eslinger, P. J., Flaherty-Craig, C. V., & Benton, A. L. (2004). Developmental outcomes after early prefrontal cortex damage. *Brain and Cognition, 55*, 84-103.

Eslinger, P. J., Parkinson, K., & Shamay, S. G. (2002). Empathy and social-emotional factors in recovery from stroke. *Current Opinion in Neurology, 15,* 91-97.

Eugene, F., Levesque, J., Mensour, B., Leroux, J. M., Beaudoin, G., Bourgouin, P., et al. (2003). The impact of individual differences on the neural circuitry underlying sadness. *NeuroImage, 19,* 354-364.

Evans, E. G. (1984). Hostility and sex guilt: Perceptions of self and others as a function of gender and sex role orientation. *Sex Roles, 10,* 207-215.

Faymonville, M. E., Laureys, S., Degueldre, C., DelFiore, G., Luxen, A., Franck, G., et al. (2000). Neural mechanisms of antinociceptive effects of hypnosis. *Anesthesiology, 92,* 1257-1267.

Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology, 69*, 1013-1027.

Ferguson, T. J., & Crowley, S. L. (1997). Gender differences in the organization of guilt and shame. *Sex Roles, 37*, 19-44.

Ferguson, T. J., & Stegge, H. (1995). Emotional states and traits in children: The case of guilt and shame. In J. P. Tangney & K. W. Fischer (Eds.), *Self-conscious emotions: The psychology of shame, guilt, embarrassment, and pride* (pp. 174-197). New York: Guilford.

Ferguson, T. J., Stegge, H., Eyre, H. L., Vollmer, R., & Ashbaker, M. (2000). Context effects and the (mal)adaptive nature of guilt and shame in children. *Genetic, Social, and General Psychology Monographs, 126*, 319-345.

Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford, CA: Stanford University Press.

Finger, E. C., Marsh, A. A., Kamel, N., Mitchell, D. G. V., & Blair, J. R. (2006). Caught in the act: The impact of audience on the neural response to morally and socially inappropriate behaviour. *NeuroImage, 33*, 414-421.

Fink, G. R., Markowitsch, H. J., Reinkemeier, M., Bruckbauer, T., Kessler, J., & Heiss, W. D. (1996). Cerebral representation of one's own past: Neural networks involved in autobiographical memory. *The Journal of Neuroscience, 16*, 4275-4282.

Fischer, A. H., & Manstead, A. R. (2000). The relation between gender and emotions in different cultures. In A. H. Fischer (Ed.), *Gender and emotion: Social psychology perspectives* (pp. 71-94). New York: Cambridge University Press.

Fischer, A. H., & van Kleef, G. A. (2010). Where have all the people gone? A plea for including social interaction in emotion research. *Emotion Review, 2*, 208-211.

Fischer, D. G., & Fick, C. (1993). Measuring social desirability: Short forms of the Marlowe–Crowne Social Desirability scale. *Educational and Psychological Measurement, 53*, 417-424.

Fisher, J. P., Young, C. N., & Fadel, P. J. (2009). Central sympathetic overactivity: Maladies and mechanisms. *Autonomic Neuroscience: Basic and Clinical, 148*, 5-15.

Fitzgerald, D. A., Posse, S., Moore, G. J., Tancer, M. E., Nathan, P. J., & Phan, K. L. (2004). Neural correlates of internally-generated disgust via autobiographical recall: A functional magnetic resonance imaging investigation. *Neuroscience Letters, 370*, 91-96.

Fletcher, P. C., Frith, C. D., Baker, S. C., Shallice, T., Frackowiak, R. S., & Dolan, R. J. (1995). The mind's eye - precuneus activation in memory-related imagery. *NeuroImage, 2*, 195-200.

Fletcher, P. C., Happe, F., Frith, U., Baker, S. C., Dolan, R. J., Frackowiak, R. S., et al. (1995).

Other minds in the brain: A functional imaging study of "theory of mind" in story comprehension. *Cognition, 57*, 109-128.

Fontaine, J. R. J., Luyten, P., de Boeck, P., Corveleyn, J., Fernandez, M., Herrera, D., et al. (2006). Untying the gordian knot of guilt and shame. *Journal of Cross-Cultural Psychology*, *37*, 273-292.

Fowles, D. C. (1980). The three arousal model: Implications of Gray's two-factor learning theory for heart rate, electrodermal activity, and psychopathy. *Psychophysiology, 17*, 87-104.

Fowles, D. C. (1988). Psychophysiology and psychopathology: A motivational approach. *Psychophysiology, 25*, 373-391.

Fowles, D. C. (2000). Electrodermal hyporeactivity and antisocial behavior: Does anxiety mediate the relationship? *Journal of Affective Disorders, 61*, 177-189.

Fowles, D. C., Fisher, A. E., & Tranel, D. (1982). The heart beats to reward: The effect of monetary incentive on heart rate. *Psychophysiology*, *19*, 506-513.

Fowles, D. C., Kochanska, G., & Murray, K. (2000). Electrodermal activity and temperament in preschool children. *Psychophysiology, 37*, 777-787.

Frazier, T. W., Strauss, M. E., & Steinhauer, S. R. (2004). Respiratory sinus arrhythmia as an index of emotional response in young adults. *Psychophysiology, 41*, 75-83.

Fredrickson, B. L. (2001). The role of positive emotions in positive psychology. The broaden-and-build theory of positive emotions. *American Psychologist: Special Issue, 56*, 218-226.

Fredrickson, B. L., & Levenson, R. W. (1998). Positive emotions speed recovery from the cardiovascular sequelae of negative emotions. *Cognition and Emotion, 12*, 191-220.

Freud, S. (1961a). *Civilization and its discontents* (J. Strachey, Trans.). Oxford, England: Hogarth Press. (Original work published 1930).

Freud, S. (1961b). *The id and the ego* (J. Strachey, Trans.). London: Hogarth Press. (Original work published 1923)

Freud, S. (1965). *The psychopathology of everyday life*. New York: Norton. (Original work published 1901).

Friedman, B. H., & Thayer, J. F. (1998). Autonomic balance revisited: Panic anxiety and heart rate variability. *Journal of Psychosomatic Research, 44*, 133-151.

Frijda, N. H. (1986). *The emotions*. Cambridge, England: Cambridge University Press.

Frijda, N. H. (1989). Aesthetic emotions and reality. *American Psychologist, 44*, 1546-1547.

Frijda, N. H. (1994). Universal antecedents exist, and are interesting. In P. Ekman & R. J.

Davidson (Eds.), *The nature of emotion: Fundamental questions* (pp. 155-162). New York: Oxford University Press.

Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J. P., Frith, C. D., & Frackowiak, R. S. J. (1994). Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping*, *2*, 189-210.

Frith, C. D., & Frith, U. (1999). Interacting minds-A biological basis. *Science, 286*, 1692-1695.

Frith, U., & Frith, C. D. (2003). Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences, 358*, 459-473.

Frodi, A. (1976). Experiential and physiological processes mediating sex differences in behavioral aggression. *Goteberg Psychological Reports*, *6*, 18-35.

Funk, C. M., & Gazzaniga, M. S. (2009). The functional brain architecture of human morality. *Current Opinion in Neurobiology*, *19*, 678-681.

Gaertner, S. L., & Dovidio, J. F. (1977). The subtlety of white racism, arousal, and helping behavior. *Journal of Personality and Social Psychology*, *35*, 691-707.

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

Gallagher, H. L., Jack, A. I., Roepstorff, A., & Frith, C. D. (2002). Imaging the intentional stance in a competitive game. *NeuroImage*, *16*, 814-821.

Gallese, V., Keysers, C., & Rizzolatti, G. (2004). A unifying view of the basis of social cognition. *Trends in Cognitive Sciences*, *8*, 396-403.

Gallup, G. G. (1982). Self-awareness and the emergence of mind in primates. *American Journal of Primatology*, *2*, 237-248.

Gambaro, S., & Rabin, A. I. (1969). Diastolic blood pressure responses following direct and displaced aggression after anger arousal in high- and low-guilt subjects. *Journal of Personality and Social Psychology, 12*, 87-94.

García-Larrea, L., Peyron, R., Mertens, P., Gregoire, M. C., Lavenne, F., Le Bars, D., et al. (1999). Electrical stimulation of motor cortex for pain control: A combined PET-scan and electrophysiological study. *Pain*, *83*, 259-273.

Gard, M. G., & Kring, A. M. (2007). Sex differences in the time course of emotion. *Emotion*, *7*, 429-437.

Garmezy, N. (1985). Stress-resistant children: The search for protective factors. In J. Stevenson (Ed.), *Recent research in developmental psychopathology* (pp. 213-233). Oxford: Pergamon Press.

Garrada, M. E., Connel, J., & Taylor, D. C. (1991). Psychophysiologic anomalies in children with emotional and conduct disorders. *Psychological Medicine, 21*, 947-957.

Garrett, A. S., & Maddock, R. J. (2001). Time course of the subjective emotional response to aversive pictures: Relevance to fMRI studies. *Psychiatry Research, 108*, 39-48.

Gehring, W. J., & Fencsik, D. E. (2001). Functions of the medial frontal cortex in the processing of conflict and errors. *The Journal of Neuroscience, 21*, 9430-9437.

Gehring, W. J., Goss, B., Coles, M. G. H., Meyer, D. E., & Donchin, E. (1993). A neural system for error detection and compensation. *Psychological Science 4*, 385-390.

Genovese, C. R., Lazar, N. A., & Nichols, T. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *NeuroImage, 15*, 870-878.

George, M. S., Ketter, T. A., Parekh, P. I., Herscovitch, P., & Post, R. M. (1996). Gender differences in regional cerebral blood flow during transient self-induced sadness or happiness. *Biological Psychiatry, 40*, 859-871.

George, M. S., Ketter, T. A., Parekh, P. I., Horwitz, B., Herscovitch, P., & Post, R. M. (1995). Brain activity during transient sadness and happiness in healthy women. *American Journal of Psychiatry, 152*, 341-351.

Gianaros, P. J., Van Der Veen, F. M., & Jennings, J. R. (2004). Regional cerebral blood flow correlates with heart period and high-frequency heart period variability during working-memory tasks: Implications for the cortical and subcortical regulation of cardiac autonomic activity. *Psychophysiology, 41*, 521-530.

Glover, G. H. (1999). 3D z-shim method for reduction of susceptibility effects in BOLD fMRI. *Magnetic Resonance in Medicine, 42*, 290-299.

Gobbini, M. I., Leibenluft, E., Santiago, N., & Haxby, J. V. (2004). Social and emotional attachment in the neural representation of faces. *NeuroImage, 22*, 1628-1635.

Goedhart, A. D., Kupper, N., Willemsen, G., Boomsma, D. I., & de Geus, E. J. (2006). Temporal stability of ambulatory stroke volume and cardiac output measured by impedance cardiography. *Biological Psychology, 72*, 110-117.

Goedhart, A. D., van der Sluis, S., Houtveen, J. H., Willemsen, G., & de Geus, E. J. C. (2007). Comparison of time and frequency domain measures of RSA in ambulatory recordings. *Psychophysiology, 44*, 203-215.

Goedhart, A. D., Willemsen, G., Houtveen, J. H., Boomsma, D. I., & de Geus, E. J. (2008). Comparing low frequency heart rate variability and preejection period: Two sides of a different coin. *Psychophysiology, 45*, 1086-1090.

Goldman-Rakic, P. S., Selemon, L. D., & Schwartz, M. L. (1984). Dual pathways connecting the dorsolateral prefrontal cortex with the hippocampal formation and parahippocampal cortex in the rhesus monkey. *Neuroscience, 12,* 719-743.

Gomez, R., & McLaren, S. (1997). The effects of reward and punishment on response disinhibition, moods, heart rate and skin conductance level during instrumental learning. *Personality and Individual Differences, 23*, 305-316.

Gotlib, I. H. (1984). Depression and general psychopathology in university students. *Journal of Abnormal Psychology, 93,* 19-30.

Grafman, J., Schwab, K., Warden, D., Pridgen, A., Brown, H. R., & Salazar, A. M. (1996). Frontal lobe injuries, violence, and aggression: A report of the Vietnam Head Injury Study. *Neurology, 46*, 1231-1238.

Graham, F. K. (1979). Distinguishing among orienting, defense, and startle reflexes. In H. D. Kimmel, E. H. v. Olst & J. F. Orlebeke (Eds.), *The orienting reflex in humans. An international conference sponsored by the Scientific Affairs Division of the North Atlantic Treaty Organization* (pp. 137-167). Hillsdale, NJ: Erlbaum.

Gray, J. A. (1972). The psychophysiological basis of introversion-extraversion: A modification of Eysenck's theory. In V. D. Nebylitsyn & J. A. Gray (Eds.), *The Biological Bases of Individual Behaviour* (pp. 182-205). San Diego, CA: Academic Press.

Gray, J. A. (1975). *Elements of a two-process theory of learning*. London: Academic Press.

Gray, J. A. (1981). A critique of Eysenck's theory of personality. In H. J. Eysenck (Ed.), *A Model for Personality* (pp. 246-276). Berlin: Springer-Verlag.

Gray, J. A. (1982). *The neuropsychology of anxiety: An enquiry into the functions of the septo-hippocampal system*. London: Oxford University Press.

Gray, J. A. (1987a). Perspectives on anxiety and impulsivity: A commentary. *Journal of Research in Personality, 21*, 493-509.

Gray, J. A. (1987b). *The psychology of fear and stress*. New York: Cambridge University Press.

Gray, J. A., & McNaughton, N. (2000). *The neuropsychology of anxiety* (2nd ed.). New York: Oxford University Press.

Gray, J. A., Owen, S., Davis, N., & Tsaltas, E. (1983). Psychological and physiolgical relations between anxiety and impulsivity. In M. Zuckerman (Ed.), *Biological bases of sensation seeking, impulsivity, and anxiety* (pp. 181-217). Hillsdale, NJ: Erlbaum.

Greene, J. D., & Haidt, J. (2002). How (and where) does moral judgment work? *Trends in Cognitive Sciences, 6,* 517-523.

Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron, 44*, 389-400.

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science, 293*, 2105-2108.

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

Greenwald, A. G., Banaji, M. R., Rudman, L. A., Farnham, S. D., Nosek, B. A., & Mellott, D. S. (2002). A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological Review, 109*, 3-25.

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

Greenwald, A. G., & Nosek, B. A. (2001). Health of the Implicit Association Test at age 3. *Zeitschrift für Experimentelle Psychologie, 48*, 85-93.

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

Greenwald, M. K., Cook, E. W., & Lang, P. J. (1989). Affective judgment and psychophysiological response: Dimensional covariation in the evaluation of pictorial stimuli. *Journal of Psychophysiology, 3*.

Grèzes, J., Berthoz, S., & Passingham, R. E. (2006). Amygdala activation when one is the target of deceit: Did he lie to you or to someone else? *NeuroImage, 30*, 601-608.

Grèzes, J., Frith, C., & Passingham, R. E. (2004). Brain mechanisms for inferring deceit in the actions of others. *The Journal of Neuroscience, 24*, 5500-5505.

Gross, J. J. (1998). The emerging field of emotion regulation: An integrative review. *Review of General Psychology, 2*, 271-299.

Gross, J. J. (2010). The future's so bright, I gotta wear shades. *Emotion Review, 2*, 212-216.

Gross, J. J., & John, O. P. (1998). Mapping the domain of expressivity: Multimethod evidence for a hierarchical model. *Journal of Personality and Social Psychology, 74*, 170-191.

Gross, J. J., & Levenson, R. W. (1997). Hiding feelings: The acute effects of inhibiting negative and positive emotion. *Journal of Abnormal Psychology, 106*, 95-103.

Grossman, M., & Wood, W. (1993). Sex differences in intensity of emotional experience: A social role interpretation. *Journal of Personality and Social Psychology, 65*, 1010-1022.

Grossman, P., & Taylor, E. W. (2007). Toward understanding respiratory sinus arrhythmia: Relations to cardiac vagal tone, evolution and biobehavioral functions. *Biological*

Psychology, 74, 263-285.

Grossman, P., Wilhelm, F. H., & Spoerle, M. (2004). Respiratory sinus arrhythmia, cardiac vagal control, and daily activity. *American Journal of Physiology: Heart and Circulatory Physiology, 287*, H728-H734.

Gusnard, D. A., Akbudak, E., Shulman, G. L., & Raichle, M. E. (2001). Medial prefrontal cortex and self-referential mental activity: Relation to a default mode of brain function. *Proceedings of the National Academy of Sciences of the United States of America, 98*, 4259-4264.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108*, 814-834.

Haidt, J. (2003). The moral emotions. In R. J. Davidson, K. R. Scherer & H. H. Goldsmith (Eds.), *Handbook of affective sciences* (pp. 852-870). New York: Oxford University Press.

Haidt, J. (2007). The new synthesis in moral psychology. *Science, 316*, 998-1002.

Hamann, S., & Canli, T. (2004). Individual differences in emotion processing. *Current Opinion in Neurobiology, 14*, 233-238.

Hamann, S., & Mao, H. (2002). Positive and negative emotional verbal stimuli elicit activity in the left amygdala. *Neuroreport, 13*, 15-19.

Hamann, S. B., Ely, T. D., Hoffman, J. M., & Kilts, C. D. (2002). Ecstasy and agony: Activation of the human amygdala in positive and negative emotion. *Psychological Science, 13*, 135-141.

Hanamori, T. (2005). Fluctuations of the spontaneous discharge in the posterior insular cortex neurons are associated with changes in the cardiovascular system in rats. *Brain Research, 1042*, 82-91.

Harder, D. W. (1995). Shame and guilt assessment and relationships of shame and guilt proneness to psychopathology. In J. P. Tangney & K. W. Fischer (Eds.), *Self-conscious emotions: The psychology of shame, guilt, embarrassment, and pride* (pp. 368-392). New York: Guilford Press.

Harenski, C. L., & Hamann, S. (2006). Neural correlates of regulating negative emotions related to moral violations. *NeuroImage, 30*, 313-324.

Harmon-Jones, E. (2000). Cognitive Dissonance and Experienced Negative Affect: Evidence that Dissonance Increases Experienced Negative Affect Even in the Absence of Aversive Consequences. *Personality and Social Psychology Bulletin, 26*, 1490-1501.

Harmon-Jones, E. (2003). Anger and the behavioral approach system. *Personality and Individual Differences, 35*, 995-1005.

Harmon-Jones, E., Amodio, D. M., & Zinner, L. R. (2007). Social psychological methods in emotion elicitation. In J. A. Coan & J. J. B. Allen (Eds.), *Handbook of emotion elicitation and assessment* (pp. 91-105). New York: Oxford University Press.

Harmon-Jones, E., Gable, P. A., & Peterson, C. K. (2009). The role of asymmetric frontal cortical activity in emotion-related phenomena: A review and update. *Biological Psychology*.

Harmon-Jones, E., & Peterson, C. K. (2009). Supine body position reduces neural response to anger evocation. *Psychological Science 20*, 1209-1210.

Harmon-Jones, E., Peterson, C. K., & Harris, C. R. (2009). Jealousy: Novel methods and neural correlates. *Emotion, 9*, 113-117.

Harmon-Jones, E., & Sigelman, J. (2001). State anger and prefrontal brain activity: evidence that insult-related relative left-prefrontal activation is associated with experienced anger and aggression. *Journal of Personality and Social Psychology, 80*, 797-803.

Harrison, N. A., Gray, M. A., Gianaros, P. J., & Critchley, H. D. (2010). The embodiment of emotional feelings in the brain. *The Journal of Neuroscience, 30*, 12878-12884.

Hasui, C., Kitamura, T., Tamaki, A., Takahashi, M., Masuda, S., & Ozeki, N. (2009). The Test of Self-Conscious Affect-3 in Japanese university students. *The Journal of Nervous and Mental Disease*, *197*, 458-460.

Hatch, J. P., Klatt, K., Porges, S. W., Schroeder-Jasheway, L., & Supik, J. D. (1986). The relation between rhythmic cardiovascular variability and reactivity to orthostatic, cognitive, and cold pressor stress. *Psychophysiology*, *23*, 48-56.

Hayano, J., Sakakibara, Y., Yamada, A., Yamada, M., Mukai, S., Fujinami, T., et al. (1991). Accuracy of assessment of cardiac vagal tone by heart rate variability in normal subjects. *American Journal of Cardiology, 67*, 199-204.

Heekeren, H. R., Wartenburger, I., Schmidt, H., Schwintowski, H. P., & Villringer, A. (2003). An fMRI study of simple ethical decision-making. *Neuroreport, 14*, 1215-1219.

Heerey, E. A., Keltner, D., & Capps, L. M. (2003). Making sense of self-conscious emotion: Linking theory of mind and emotion in children with autism. *Emotion, 3*, 394-400.

Heider, F. (1958). *The psychology of interpersonal relations*. New York: John Wiley.

Heims, H. C., Critchley, H. D., Dolan, R., Mathias, C. J., & Cipolotti, L. (2004). Social and motivational functioning is not critically dependent on feedback of autonomic responses: Neuropsychological evidence from patients with pure autonomic failure. *Neuropsychologia, 42*, 1979-1988.

Henning, K. R., & Frueh, B. C. (1997). Combat guilt and its relationship to PTSD symptoms. *Journal of Clinical Psychology, 53*, 801-808.

Heponiemi, T., Keltikangas-Järvinen, L., Kettunen, J., Puttonen, S., & Ravaja, N. (2004). BIS-BAS sensitivity and cardiac autonomic stress profiles. *Psychophysiology, 41*, 37-45.

Heponiemi, T., Keltikangas-Järvinen, L., Puttonen, S., & Ravaja, N. (2003). BIS/BAS sensitivity and self-rated affects during experimentally induced stress. *Personality and Individual Differences, 34*, 943-957.

Herek, G. M. (2000). Sexual prejudice and gender: Do heterosexuals' attitudes toward lesbians and gay men differ? *Journal of Social Issues, 56*, 251-266.

Herrald, M. M., & Tomaka, J. (2002). Patterns of emotion-specific appraisal, coping, and cardiovascular reactivity during an ongoing emotional episode. *Journal of Personality and Social Psychology, 83*, 434 - 450.

Higgins, E. T. (1987). Self-discrepancy theory: A theory relating self and affect. *Psychological Review, 94*, 319-340.

Higgins, E. T. (1999). When do self-discrepancies have specific relations to emotions? The second-generation question of Tangney, Niedenthal, Covert, and Barlow (1998). *Journal of Personality and Social Psychology, 77*, 1313-1317.

Higgins, E. T., Bond, R. N., Klein, R., & Strauman, T. J. (1986). Self-discrepancies and emotional vulnerability: How magnitude, accessibility, and type of discrepancy influence affect. *Journal of Personality and Social Psychology, 51*, 5-15.

Hoffman, M. L. (1975). Sex differences in moral internalization and values. *Journal of Personality and Social Psychology, 32*, 720-729.

Hoffman, M. L. (1982). Development of prosocial motivation: Empathy and guilt. In N. Eisenberg (Ed.), *The Development of Prosocial Behaviour* (pp. 281-313). New York: Academic.

Hoffman, M. L. (1998). Varieties of empathy-based guilt. In J. Bybee (Ed.), *Guilt and children* (Vol. 4, pp. 91-112). New York: Academic.

Hogan, R. (1973). Moral conduct and moral character: A psychological perspective. *Psychological Bulletin, 79*, 217-232.

Holmes, A. P., & Friston, K. J. (1998). Generalisability, random effects and population inference. *Neuroimage, 7*, S754.

Holroyd, C. B., Nieuwenhuis, S., Yeung, N., Nystrom, L., Mars, R. B., Coles, M. G. H., et al. (2004). Dorsal anterior cingulate cortex shows fMRI response to internal and external error signals. *Nature Neuroscience, 7*, 497-498.

Houston, D. A. (1990). Empathy and the self: Cognitive and emotional influences on the evaluation of negative affect in others. *Journal of Personality and Social Psychology, 59*, 859-868.

Houtveen, J. H., Rietveld, S., & de Geus, E. J. (2002). Contribution of tonic vagal modulation of heart rate, central respiratory drive, respiratory depth, and respiratory frequency to respiratory sinus arrhythmia during mental stress and physical exercise. *Psychophysiology*, *39*, 427-436.

Hsu, M., Anen, C., & Quartz, S. R. (2008). The right and the good: Distributive justice and neural encoding of equity and efficiency. *Science*, *320*, 1092-1095.

Hubert, W., & de Jong-Meyer, R. (1991). Autonomic, neuroendocrine, and subjective responses to emotion-inducing film stimuli. *International Journal of Psychophysiology, 11*, 131-140.

Hume, D. (1960). *An enquiry concerning the principles of morals*. La Salle, IL: Open Court. (Original work published 1777).

Hume, D. (1984). *A treatise of human nature*. New York: Penguin Classics. (Original work published 1739).

Humphrey (1986). *The inner eye*. London: Faber & Faber.

Immordino-Yang, M. H., McColl, A., Damasio, H., & Damasio, A. (2009). Neural correlates of admiration and compassion. *Proceedings of the National Academy of Sciences of the United States of America*, *106*, 8021-8026.

Insel, T. R., & Young, L. J. (2001). The neurobiology of attachment. *Nature Reviews Neuroscience*, *2*, 129-136.

Izard, C. E. (1977). *Human emotions*. New York: Plenum Press.

Izard, C. E. (1991). *The psychology of emotions*. New York: Plenum Press.

Izard, C. E. (1992). Basic emotions, relations among emotions, and emotion-cognition relations. *Psychological Review*, *99*, 561-565.

Izard, C. E., & Ackerman, B. P. (2000). Motivational, organizational, and regulatory functions of discrete emotions. In M. Lewis & J. M. Haviland-Jones (Eds.), *Handbook of emotions* (2nd ed.). New York: Guilford Press.

Jackson, D. C., Mueller, C. J., Dolski, I., Dalton, K. M., Nitschke, J. B., Urry, H. L., et al. (2003). Now you feel it, now you don't: Frontal brain electrical asymmetry and individual differences in emotion regulation. *Psychological Science*, *14*, 612-617.

James, W. (1884). What is an emotion? *Mind*, *9*, 188-205.

James, W. (1892). *Principles of psychology*. New York: Henry Holt and Company.

James, W. (1894). The physical basis of emotion. *Psychological Review*, 1, 516-529. Reprinted 1994. *Psychological Review*, *101*, 205-210.

Janoff-Bulman, R., Sheikh, S., & Hepp, S. (2009). Proscriptive versus prescriptive morality: Two faces of moral regulation. *Journal of Personality and Social Psychology, 96*, 521-537.

Jennings, J. R., & van der Molen, M. W. (2002). Cardiac timing and the central regulation of action. *Psychological Research 66*, 337-349.

Johansen-Berg, H., Gutman, D. A., Behrens, T. E., Matthews, P. M., Rushworth, M. F., Katz, E., et al. (2008). Anatomical connectivity of the subgenual cingulate region targeted with deep brain stimulation for treatment-resistant depression. *Cerebral Cortex, 18*, 1374-1383.

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

Jones, P. P., Spraul, M., Matt, K. S., Seals, D. R., Skinner, J. S., & Ravussin, E. (1996). Gender does not influence sympathetic neural reactivity to stress in healthy humans. *American Journal of Physiology - Heart and Circulatory Physiology, 270*, H350-H357.

Jorm, A. F., Christensen, H., Henderson, A. S., Jacomb, P. A., Korten, A. E., & Rodgers, B. (1999). Using the BIS/BAS scales to measure behavioural inhibition and behavioural activation: Factor structure, validity and norms in a large community sample. *Personality and Individual Differences, 26*, 49-58.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience, 17*, 4302-4311.

Katkin, E. S. (1965). Relationship between manifest anxiety and two indices of autonomic response to stress. *Journal of Personality and Social Psychology, 2*, 324-333.

Katkin, E. S., Wiens, S., & Öhman, A. (2001). Nonconscious fear conditioning, visceral perception, and the development of gut feelings. *Psychological Science, 12*, 366-370.

Katz, I., Wackenhut, J., & Glass, D. C. (1986). An ambivalence-amplification theory of behavior towards the stigmatized. In S. Worchel & W. G. Austen (Eds.), *Psychology of intergroup relations* (2nd ed., pp. 103-117). Chicago: Nelson-Hall.

Kédia, G., Berthoz, S., Wessa, M., Hilton, D., & Martinot, J.-L. (2008). An agent harms a victim: A functional magnetic resonance imaging study on specific moral emotions. *Journal of Cognitive Neuroscience, 20*, 1788-1798.

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

Kelman, H. C. (1958). Compliance, identification and internalization, three processes of attitude change. *Journal of Conflict Resolution, 2*, 51-60.

Keltner, D. (1995). Signs of appeasement: Evidence for the distinct displays of embarrassment, amusement, and shame. *Journal of Personality and Social Psychology, 68*, 441-454.

Keltner, D., & Buswell, B. N. (1997). Embarrassment: Its distinct form and appeasement functions. *Psychological Bulletin, 122*, 250-270.

Keltner, D., & Haidt, J. (1999). Social functions of emotions at four levels of analysis. *Cognition and Emotion, 13*, 505-521.

Keltner, D., & Harker, L. (1998). The forms and functions of the nonverbal signal of shame. In P. Gilbert & B. Andrews (Eds.), *Shame: Interpersonal behavior, psychopathology, and culture* (pp. 78-98). Oxford: Oxford University Press.

Keltner, D. A. C. (2000). Saving face for Darwin: the functions and uses of embarrassment. *Current Directions in Psychological Science, 9*, 187-192.

Kerns, J. G., Cohen, J. D., MacDonald, A. W., 3rd, Cho, R. Y., Stenger, V. A., & Carter, C. S. (2004). Anterior cingulate conflict monitoring and adjustments in control. *Science, 303*, 1023-1026.

Khalfa, S., Isabelle, P., Jean-Pierre, B., & Manon, R. (2002). Event-related skin conductance responses to musical emotions in humans. *Neuroscience Letters, 328*, 145-149.

Kiehl, K. A., Smith, A. M., Hare, R. D., Mendrek, A., Forster, B. B., Brink, J., et al. (2001). Limbic abnormalities in affective processing by criminal psychopaths as revealed by functional magnetic resonance imaging. *Biological Psychiatry, 50*, 677-684.

Kim, J., Talbot, N. L., & Cicchetti, D. (2009). Childhood abuse and current interpersonal conflict: The role of shame. *Child Abuse and Neglect, 33*, 362-371.

Kimbrell, T. A., George, M. S., Parekh, P. I., Ketter, T. A., Podell, D. M., Danielson, A. L., et al. (1999). Regional brain activity during transient self-induced anxiety and anger in healthy adults. *Biological Psychiatry, 46*, 454-465.

Kinomura, S., Kawashima, R., Yamada, K., Ono, S., Itoh, M., Yoshioka, S., et al. (1994). Functional anatomy of taste perception in the human brain studied with positron emission tomography. *Brain Research 659*, 263-266.

Kircher, T. T., Senior, C., Phillips, M. L., Benson, P. J., Bullmore, E. T., Brammer, M., et al. (2000). Towards a functional neuroanatomy of self processing: Effects of faces and words. *Brain Research. Cognitive Brain Research, 10*, 133-144.

Kjaer, T. W., Nowak, M., & Lou, H. C. (2002). Reflective self-awareness and conscious states: PET evidence for a common midline parietofrontal core. *NeuroImage, 17*, 1080-1086.

Knyazev, G. G., Slobodskaya, H. R., & Wilson, G. D. (2002). Psychophysiological correlates of behavioural inhibition and activation. *Personality and Individual Differences, 33*, 647-660.

Kochanska, G., Gross, J. N., Lin, M.-H., & Nichols, K. E. (2002). Guilt in young children: Development, determinants, and relations with a broader system of standards. *Child Development*, *73*, 461-482.

Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M., et al. (2007). Damage to the prefrontal cortex increases utilitarian moral judgements. *Nature*, *446*, 908-911.

Kohlberg, L. (1969). Stage and sequence: The cognitive-developmental approach to socialization. In D. A. Goslin (Ed.), *Handbook of of socialization theory and research* (pp. 347-480). Chicago: Rand McNally.

Koski, L., & Paus, T. (2000). Functional connectivity of the anterior cingulate cortex within the human frontal lobe: A brain-mapping meta-analysis. *Experimental Brain Research, 133*, 55-65.

Krajbich, I., Adolphs, R., Tranel, D., Denburg, N. L., & Camerer, C. F. (2009). Economic games quantify diminished sense of guilt in patients with damage to the prefrontal cortex. *The Journal of Neuroscience, 29*, 2188-2192.

Kreibig, S. D., Wilhelm, F. H., Roth, W. T., & Gross, J. J. (2007). Cardiovascular, electrodermal, and respiratory response patterns to fear- and sadness-inducing films. *Psychophysiology, 44*, 787-806.

Kring, A. M., & Bachorowski, J.-A. (1999). Emotions and psychopathology. *Cognition & Emotion, 13*, 575-599.

Kring, A. M., & Gordon, A. H. (1998). Sex differences in emotion: Expression, experience, and physiology. *Journal of Personality and Social Psychology, 74*, 686-703.

Kringelbach, M. L., & Rolls, E. T. (2003). Neural correlates of rapid reversal learning in a simple model of human social interaction. *NeuroImage, 20*, 1371-1383.

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.

Krolak-Salmon, P., Henaff, M. A., Isnard, J., Tallon-Baudry, C., Guenot, M., Vighetto, A., et al. (2003). An attention modulated response to disgust in human ventral anterior insula. *Annals of Neurology*, *53*, 446-453.

Kroll, J., & Egan, E. (2004). Psychiatry, moral worry, and the moral emotions. *Journal of Psychiatric Practice, 10*, 352-360.

Kubany, E. S., & Watson, S. B. (2003). Guilt: Elaboration of a multidimensional model. *The Psychological Record*, *53*, 51-90.

Kugler, K., & Jones, W. H. (1992). On conceptualizing and assessing guilt. *Journal of*

Personality and Social Psychology, 62, 318-327.

Kupper, N., Willemsen, G., Boomsma, D. I., & de Geus, E. J. (2006). Heritability of indices for cardiac contractility in ambulatory recordings. *Journal of Cardiovascular Electrophysiology, 17*, 877-883.

Lagattuta, K. H., & Thompson, R. A. (2007). The development of self-conscious emotions: Cognitive processes and social influences. In J. L. Tracy, R. W. Robins & J. P. Tangney (Eds.), *The self-conscious emotions: Theory and research* (pp. 91-113). New York: Guilford.

Lamb, R. (1983). Guilt, shame and morality. *Philosophy and Phenomenological Research, 43*, 329-346.

Lamm, C., Batson, C. D., & Decety, J. (2007). The neural substrate of human empathy: Effects of perspective-taking and cognitive appraisal. *Journal of Cognitive Neuroscience, 19*, 42-58.

Lamm, C., Decety, J., & Singer, T. (2011). Meta-analytic evidence for common and distinct neural networks associated with directly experienced pain and empathy for pain. *NeuroImage, 54*, 2492-2502.

Lane, R. D., Fink, G. R., Chau, P. M. L., & Dolan, R. J. (1997). Neural activation during selective attention to subjective emotional responses. *NeuroReport, 8*, 3969-3972.

Lane, R. D., McRae, K., Reiman, E. M., Chen, K., Ahern, G. L., & Thayer, J. F. (2009). Neural correlates of heart rate variability during emotion. *NeuroImage, 44*, 213-222.

Lane, R. D., Reiman, E. M., Ahern, G. L., Schwartz, G. E., & Davidson, R. J. (1997). Neuroanatomical correlates of happiness, sadness, and disgust. *American Journal of Psychiatry, 154*, 926-933.

Lane, R. D., Reiman, E. M., Axelrod, B., Yun, L.-S., Holmes, A., & Schwartz, G. E. (1998). Neural correlates of levels of emotional awareness: Evidence of an interaction between emotion and attention in the anterior cingulate cortex. *Journal of Cognitive Neuroscience, 10*, 525.

Lang, P. J. (1980). Behavioral treatment and bio-behavioral assessment: Computer applications. In J. B. Sidowski, J. H. Johnson & T. A. Williams (Eds.), *Technology in mental health care delivery systems* (pp. 119-137). Norwood, NJ: Ablex Publishing.

Lang, P. J. (2010). Emotion and motivation: Toward consensus definitions and a common research purpose. *Emotion Review, 2*, 229-233.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1990). Emotion, attention, and the startle reflex. *Psychological Review, 97*, 377-395.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1997). Motivated attention: Affect, activation, and action. In P. J. Lang, R. F. Simons & M. T. Balaban (Eds.), *Attention and orienting: Sensory and motivational processes* (pp. 97-135). Hillsdale, NJ: Erlbaum

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1999). *International affective picture system (IAPS): Technical manual and affective ratings*. University of Florida, Center for Research in Psychophysiology, Gainesville.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2005). *International affective picture system (IAPS): Instruction manual and affective picture ratings*. Technical Report A-6, The Centre for Research in psychophysiology, University of Florida.

Lang, P. J., Greenwald, M. K., Bradley, M. M., & Hamm, A. O. (1993). Looking at pictures: Affective, facial, visceral, and behavioral reactions. *Psychophysiology, 30*, 261-273.

Lange, C. G. (1885). The mechanism of the emotions. In B. Rand (Ed.), *The classical psychologist* (pp. 672-685). Boston: Houghton Mifflin.

Lange, K., Williams, L. M., Young, A. W., Bullmore, E. T., Brammer, M. J., Williams, S. C. R., et al. (2003). Task instructions modulate neural responses to fearful facial expressions. *Biological Psychiatry*, *53*, 226-232.

Larsen, R. J. (2002). Differential contributions of positive and negative affect to subjective well-being. In J. A. Da Silva, E. H. Matsushima & N. P. Riberio-Filho (Eds.), *Annual Meeting of the International Society for Psychophysics* (Vol. 18, pp. 186-190). Rio de Janeiro, Brazil: Editora Legis Summa Ltda.

Larsen, R. J., & Diener, E. (1987). Affect intensity as an individual difference characteristic: A review. *Journal of Research in Personality, 21*, 1-39.

Larsen, R. J., & Fredrickson, B. L. (1999). Measurment issues in emotion research. In D. Kahneman, E. Diener & N. Schwarz (Eds.), *Well-being: The foundations of hedonic psychology* (pp. 40-60). New York: Russell Sage Foundation.

Larsen, R. J., & Prizmic, Z. (2004). Affect Regulation. In R. F. Baumeister & K. D. Vohs (Eds.), *Handbook of Self-Regulation. Research, Theory and Applications* (pp. 40-61). New York: The Guilford Press.

Lazarus, R. S. (1991). *Emotion and adaptation*. New York: Oxford University Press.

Leary, M. R. (2004). Digging deeper: The fundamental nature of "self-conscious" emotions. *Psychological Inquiry, 15*, 129-131.

Leary, M. R. (2007). Motivational and emotional aspects of the self. *Annual Review of Psychology, 58*, 22.01-22.28.

Leary, M. R., & Buttermore, N. E. (2003). Evolution of the human self: Tracing the natural history of self-awareness. *Journal for the Theory of Social Behaviour*, *33*, 365-404.

Leary, M. R., Koch, E., & Hechenbleikner, N. (2001). Emotional responses to interpersonal rejection. In M. R. Leary (Ed.), *Interpersonal rejection* (pp. 145-166). New York: Oxford University Press.

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

LeDoux, J. E. (2003). The emotional brain, fear, and the amygdala. *Cellular and Molecular Neurobiology, 23*, 727-738.

Lerner, J. S., Dahl, R. E., Hariri, A. R., & Taylor, S. E. (2007). Facial expressions of emotion reveal neuroendocrine and cardiovascular stress responses. *Biological Psychiatry, 61*, 253-260.

Levenson, R. W. (1992). Autonomic nervous system differences among emotions. *Psychological Science, 3*, 23-27.

Levenson, R. W. (1999). The intrapersonal functions of emotion. *Cognition & Emotion, 13*, 481-504.

Levenson, R. W. (2003a). Autonomic specificity and emotion. In R. J. Davidson, H. H. Goldsmith & K. R. Scherer (Eds.), *Handbook of affective sciences* (pp. 212-224). New York: Oxford University Press.

Levenson, R. W. (2003b). Blood, sweat, and fears: The autonomic architecture of emotion. *Annals of the New York Academy of Sciences, 1000*, 348-366.

Levenson, R. W., Carstensen, L. L., & Gottman, J. M. (1994). Influence of age and gender on affect, physiology, and their interrelations: A study of long-term marriages. *Journal of Personality and Social Psychology, 67*, 56-68.

Levenson, R. W., & Ekman, P. (2002). Difficulty does not account for emotion-specific heart rate changes in the directed facial action task. *Psychophysiology, 39*, 397-405.

Levenson, R. W., Ekman, P., & Friesen, W. V. (1990). Voluntary facial action generates emotion-specific autonomic nervous system activity. *Psychophysiology, 27*, 363-384.

Lewis, H. B. (1971). *Shame and guilt in neurosis*. New York: International Universities Press.

Lewis, M. (1995). Embarrassment: The emotion of self-exposure and evaluation. In J. P. Tangney & K. W. Fischer (Eds.), *Self-conscious emotions: The psychology of shame, guilt, embarrassment, and pride* (pp. 198-218). New York: Guilford.

Lewis, M. (2000). Self-conscious emotions: Embarrassment, pride, shame and guilt. In M. Lewis & J. M. Haviland-Jones (Eds.), *Handbook of emotions* (2nd ed., pp. 623-636). New York: Guilford Press.

Lewis, M., Sullivan, M. W., Stanger, C., & Weiss, M. (1989). Self development and self-conscious emotions. *Child Development, 60*, 146-156.

Liberzon, I., Phan, K. L., Decker, L. R., & Taylor, S. F. (2003). Extended amygdala and emotional salience: A PET activation study of positive and negative affect. *Neuropsychopharmacology, 28*, 726-733.

Liberzon, I., Taylor, S. F., Fig, L. M., Decker, L. R., Koeppe, R. A., & Minoshima, S. (2000). Limbic activation and psychophysiologic responses to aversive visual stimuli. Interaction with cognitive task. *Neuropsychopharmacology, 23*, 508-516.

Lindsay-Hartz, J. (1984). Contrasting experiences of shame and guilt. *American Behavioral Scientist, 27*, 689-704.

Liotti, M., Mayberg, H. S., Brannan, S. K., McGinnis, S., Jerabek, P., & Fox, P. T. (2000). Differential limbic--cortical correlates of sadness and anxiety in healthy subjects: Implications for affective disorders. *Biological Psychiatry, 48*, 30-42.

Logothetis, N. K. (2003). The underpinnings of the BOLD functional magnetic resonance imaging signal. *The Journal of Neuroscience, 23*, 3963-3971.

Logothetis, N. K., & Pfeuffer, J. (2004). On the nature of the BOLD fMRI contrast mechanism. *Magnetic Resonance Imaging, 22*, 1517-1531.

Lou, H. C., Luber, B., Crupain, M., Keenan, J. P., Nowak, M., Kjaer, T. W., et al. (2004). Parietal cortex and representation of the mental self. *Proceedings of the National Academy of Sciences of the United States of America, 101*, 6827-6832.

Lu, L. H., Crosson, B., Nadeau, S. E., Heilman, K. M., Gonzalez-Rothi, L. J., Raymer, A., et al. (2002). Category-specific naming deficits for objects and actions: Semantic attribute and grammatical role hypotheses. *Neuropsychologia, 40*, 1608-1621.

Luo, Q., Nakic, M., Wheatley, T., Richell, R., Martin, A., & Blair, R. J. (2006). The neural basis of implicit moral attitude - An IAT study using event-related fMRI. *NeuroImage, 30*, 1449-1457.

Luu, P., Collins, P., & Tucker, D. M. (2000). Mood, personality, and self-monitoring: Negative affect and emotionality in relation to frontal lobe mechanisms of error monitoring. *Journal of Experimental Psychology: General, 129*, 43-60.

MacDonald, A. W., Cohen, J. D., Stenger, V. A., & Carter, C. S. (2000). Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. *Science, 288*, 1835-1838.

Maddock, R. J. (1999). The retrosplenial cortex and emotion: New insights from functional neuroimaging of the human brain. *Trends in Neurosciences, 22*, 310-316.

Mak, A. K., Hu, Z. G., Zhang, J. X., Xiao, Z. W., & Lee, T. M. (2009a). Neural correlates of regulation of positive and negative emotions: An fMRI study. *Neuroscience Letters, 457*, 101-106.

Mak, A. K., Hu, Z. G., Zhang, J. X., Xiao, Z. W., & Lee, T. M. (2009b). Sex-related differences in neural activity during emotion regulation. *Neuropsychologia, 47*, 2900-2908.

Malliani, A., Pagani, M., Lombardi, F., & Cerutti, S. (1991). Cardiovascular neural regulation explored in the frequency domain. *Circulation, 84*, 482-492.

Manstead, A. S. R. (1992). Gender differences in emotion. In A. Gale & M. W. Eysenck (Eds.), *Handbook of individual differences: Biological perspectives*. Chichester, England: Wiley.

Martinmäki, K., Rusko, H., Kooistra, L., Kettunen, J., & Saalasti, S. (2006). Intraindividual validation of heart rate variability indexes to measure vagal effects on hearts. *American Journal of Physiology: Heart and Circulatory Physiology, 290*, H640-H647.

Mascolo, M. F., & Fischer, K. W. (1995). Developmental transformations in appraisals for pride, shame, and guilt. In J. P. Tangney & K. W. Fischer (Eds.), *Self-conscious emotions: The psychology of shame, guilt, embarrassment, and pride*. New York: Guilford.

Masten, C. L., Morelli, S. A., & Eisenberger, N. I. (2011). An fMRI investigation of empathy for 'social pain' and subsequent prosocial behavior. *NeuroImage, 55*, 381-388.

Matthews, S. C., Paulus, M. P., Simmons, A. N., Nelesen, R. A., & Dimsdale, J. E. (2004). Functional subdivisions within anterior cingulate cortex and their relationship to autonomic nervous system function. *NeuroImage, 22*, 1151-1156.

Mauss, I. B., Levenson, R. W., McCarter, L., Wilhelm, F. H., & Gross, J. J. (2005). The tie that binds? Coherence among emotion experience, behavior, and physiology. *Emotion, 5*, 175-190.

Maxwell, S. E. (1980). Pairwise multiple comparisons in repeated-measures designs. *Journal of Educational Statistics, 5*, 269-287.

Mayberg, H. S., Liotti, M., Brannan, S. K., McGinnis, S., Mahurin, R. K., Jerabek, P. A., et al. (1999). Reciprocal limbic-cortical function and negative mood: Converging PET findings in depression and normal sadness. *American Journal of Psychiatry, 156*, 675-682.

Mayne, T. J. (1999). Negative affect and health: The importance of being earnest. *Cognition & Emotion, 13*, 601-635.

McCraty, R., Atkinson, M., Tiller, W. A., Rein, G., & Watkins, A. D. (1995). The effects of emotions on short-term power spectrum analysis of heart rate variability. *American Journal of Cardiology, 76*, 1089-1093.

McEwen, B. S. (2008). Central effects of stress hormones in health and disease: Understanding the protective and damaging effects of stress and stress mediators. *European Journal of Pharmacology, 583*, 174-185.

McNaughton, N., & Corr, P. J. (2004). A two-dimensional neuropsychology of defense: Fear/anxiety and defensive distance. *Neuroscience and Biobehavioral Reviews, 28,* 285-305.

Meissner, W. W. (1981). *Internalization in psychoanalysis*. New York: International Universities Press.

Mesulam, M.-M. (1985). *Principles of Behavioral Neurology*. Philadelphia: F. A. Davis.

Mesulam, M. M., & Mufson, E. J. (1982a). Insula of the old world monkey. I. Architectonics in the insulo-orbito-temporal component of the paralimbic brain. *Journal of Comparative Neurology, 212*, 1-22.

Mesulam, M. M., & Mufson, E. J. (1982b). Insula of the old world monkey. III: Efferent cortical output and comments on function. *Journal of Comparative Neurology, 212*, 38-52.

Mezzacappa, E., Tremblay, R. E., Kindlon, D., Saul, J. P., Arseneault, L., Seguin, J., et al. (1997). Anxiety, antisocial behavior, and heart rate regulation in adolescent males. *Journal of Child Psychology and Psychiatry, 38*, 457-469.

Miller, B. L., Chang, L., Mena, I., Boone, K., & Lesser, I. M. (1993). Progressive right frontotemporal degeneration: Clinical, neuropsychological and SPECT characteristics. *Dementia, 4*, 204-213.

Miller, B. L., Darby, A., Benson, D. F., Cummings, J. L., & Miller, M. H. (1997). Aggressive, socially disruptive and antisocial behaviour associated with fronto-temporal dementia. *The British Journal of Psychiatry, 170*, 150-154.

Miller, N. E. (1961). Some recent studies on conflict behavior and drugs. *American Psychologist, 16*, 12-24.

Miller, R. S. (1996). *Embarrassment: Poise and peril in everyday life*. New York: Guilford Press.

Miller, W. I. (1997). *The anatomy of disgust*. Cambridge: Harvard University Press.

Moak, J. P., Goldstein, D. S., Eldadah, B. A., Saleem, A., Holmes, C., Pechnik, S., et al. (2009). Supine low-frequency power of heart rate variability reflects baroreflex function, not cardiac sympathetic innervation. *Cleveland Clinic Journal of Medicine, 76*, S51-S59.

Moll, J., de Oliveira-Souza, R., Bramati, I. E., & Grafman, J. (2002). Functional networks in emotional moral and nonmoral social judgments. *NeuroImage, 16*, 696-703.

Moll, J., de Oliveira-Souza, R., Eslinger, P. J., Bramati, I. E., Mourao-Miranda, J., Andreiuolo, P. A., et al. (2002). The neural correlates of moral sensitivity: A functional magnetic resonance imaging investigation of basic and moral emotions. *The Journal of Neuroscience, 22*, 2730-2736.

Moll, J., de Oliveira-Souza, R., Garrido, G. J., Bramati, I. E., Caparelli-Daquer, E. M. A., Paiva, M. L. M. F., et al. (2007). The self as a moral agent: Linking the neural bases of social agency and moral sensitivity. *Social Neuroscience, 2*, 336-352.

Moll, J., de Oliveira-Souza, R., Moll, F. T., Ignacio, F. A., Bramati, I. E., Caparelli-Daquer, E. M., et al. (2005). The moral affiliations of disgust: A functional MRI study. *Cognitive Behavioral Neurology, 18*, 68-78.

Moll, J., De Oliveira-Souza, R., & Zahn, R. (2008). The neural basis of moral cognition. *Annals of the New York Academy of Sciences, 1124*, 161-180.

Moll, J., Eslinger, P. J., & Oliveira-Souza, R. (2001). Frontopolar and anterior temporal cortex activation in a moral judgment task: Preliminary functional MRI results in normal subjects. *Arquivos de Neuro-psiquiatria, 59,* 657-664.

Moll, J., Krueger, F., Zahn, R., Pardini, M., de Oliveira-Souza, R., & Grafman, J. (2006). Human fronto-mesolimbic networks guide decisions about charitable donation. *Proceedings of the National Academy of Sciences of the United States of America, 103*, 15623-15628.

Moll, J., Zahn, R., de Oliveira-Souza, R., Krueger, F., & Grafman, J. (2005). Opinion: The neural basis of human moral cognition. *Nature Reviews Neuroscience, 6*, 799-809.

Montague, P. R., Berns, G. S., Cohen, J. D., McClure, S. M., Pagnoni, G., Dhamala, M., et al. (2002). Hyperscanning: Simultaneous fMRI during linked social interactions. *NeuroImage, 16,* 1159-1164.

Montague, P. R., & Lohrenz, T. (2007). To detect and correct: Norm violations and their enforcement. *Neuron, 56,* 14-18.

Montano, N., Porta, A., Cogliati, C., Costantino, G., Tobaldini, E., Casali, K. R., et al. (2009). Heart rate variability explored in the frequency domain: A tool to investigate the link between heart and behavior. *Neuroscience and Biobehavioral Reviews, 33*, 71-80.

Monteith, M. J. (1993). Self-regulation of prejudiced responses: Implications for progress in prejudice-reduction efforts. *Journal of Personality and Social Psychology, 65,* 469-485.

Monteith, M. J., Ashburn-Nardo, L., Voils, C. I., & Czopp, A. M. (2002). Putting the brakes on prejudice: On the development and operation of cues for control. *Journal of Personality and Social Psychology, 83*, 1029-1050.

Monteith, M. J., Devine, P. G., & Zuwerink, J. R. (1993). Self-directed versus other-directed affect as a consequence of prejudice-related discrepancies. *Journal of Personality and Social Psychology, 64*, 198-210.

Monteith, M. J., Voils, C. I., & Ashburn-Nardo, L. (2001). Taking a look underground: Detecting, interpreting, and reacting to implicit racial biases. *Social Cognition, 19*, 395.

Morgan, B. E. (2006). Behavioral inhibition: A neurobiological perspective. *Current Psychiatry*

Reports, 8, 270-278.

Morris, J. S., Frith, C. D., Perrett, D. I., Rowland, D., Young, A. W., Calder, A. J., et al. (1996). A differential neural response in the human amygdala to fearful and happy facial expressions. *Nature, 383*, 812-815.

Morrison, S. F. (2001). Differential control of sympathetic outflow. *American Journal of Physiology. Regulatory, Integrative and Comparative Physiology, 281*, R683-698.

Muller, J. L., Sommer, M., Wagner, V., Lange, K., Taschler, H., Roder, C. H., et al. (2003). Abnormalities in emotion processing within cortical and subcortical regions in criminal psychopaths: Evidence from a functional magnetic resonance imaging study using pictures with emotional content. *Biological Psychiatry, 54*, 152-162.

Myrdal, G. (1944). *An American dilemma*. New York: Harper & Row.

Nagai, Y., Critchley, H. D., Featherstone, E., Trimble, M. R., & Dolan, R. J. (2004). Activity in ventromedial prefrontal cortex covaries with sympathetic skin conductance level: A physiological account of a "default mode" of brain function. *NeuroImage, 22*, 243-251.

Nater, U. M., Abbruzzese, E., Krebs, M., & Ehlert, U. (2006). Sex differences in emotional and psychophysiological responses to musical stimuli. *International Journal of Psychophysiology, 62*, 300-308.

Nelissen, R. M. A., & Zeelenberg, M. (2009). When guilt evokes self-punishment: Evidence for the existence of a Dobby Effect. *Emotion, 9*, 118-122.

Nelson, E. E., & Panksepp, J. (1998). Brain substrates of infant-mother attachment: Contributions of opioids, oxytocin, and norepinephrine. *Neuroscience and Biobehavioral Reviews, 22*, 437-452.

Newton, T. L. (2009). Cardiovascular functioning, personality, and the social world: The domain of hierarchical power. *Neuroscience and Biobehavioral Reviews, 33*, 145-159.

Nichols, S. (2002). Norms with feeling: Towards a psychological account of moral judgment. *Cognition, 84*, 221-236.

Niedenthal, P. M., Tangney, J. P., & Gavanski, I. (1994). ‘‘If only I weren’t’’ versus ‘‘If only I hadn’t’’: Distinguishing shame and guilt in counterfactual thinking. *Journal of Personality and Social Psychology, 67*, 584-595.

Nielsen, L., & Kaszniak, A. W. (2007). Conceptual, theoretical, and methodological issues in inferring subjective emotion experience: Recommendations for researchers. In J. A. Coan & J. J. B. Allen (Eds.), *Handbook of emotion elicitation and assessment*. New York: Oxford University Press.

Nieuwenhuis, S., Holroyd, C. B., Mol, N., & Coles, M. G. H. (2004). Reinforcement-related brain potentials from medial frontal cortex: Origins and functional significance.

Neuroscience and Biobehavioral Reviews, 28, 441-448.

Nieuwenhuis, S., Yeung, N., van den Wildenberg, W., & Ridderinkhof, K. R. (2003). Electrophysiological correlates of anterior cingulate function in a go/no-go task: Effects of response conflict and trial type frequency. *Cognitive, Affective and Behavioral Neuroscience, 3*, 17-26.

Nisbett, R., & Wilson, T. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*, 231-259.

Nolen-Hoeksema, S., Morrow, J., & Fredrickson, B. L. (1993). Response styles and the duration of episodes of depressed mood. *Journal of Abnormal Psychology, 102,* 20-28.

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

Northoff, G., Heinzel, A., de Greck, M., Bermpohl, F., Dobrowolny, H., & Panksepp, J. (2006). Self-referential processing in our brain--A meta-analysis of imaging studies on the self. *NeuroImage, 31*, 440-457.

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2005). Understanding and using the Implicit Association Test: II. Method variables and construct validity. *Personality and Social Psychology Bulletin, 31*, 166-180.

Nosek, B. A., Smyth, F. L., Hansen, J. J., Devos, T., Lindner, N. M., Ranganath, K. A., et al. (2007). Pervasiveness and correlates of implicit attitudes and stereotypes. *European Review of Social Psychology, 18*, 36-53.

O'Brien, R. G., & Kaiser, M. K. (1985). MANOVA method for analyzing repeated measures designs: an extensive primer. *Psychol Bull, 97*, 316-333.

O'Connor, L. E., Berry, J. W., Weiss, J., & Gilbert, P. (2002). Guilt, fear, submission, and empathy in depression. *Journal of Affective Disorders, 71*, 19-27.

O'Doherty, J., Critchley, H., Deichmann, R., & Dolan, R. J. (2003). Dissociating valence of outcome from behavioral control in human orbital and ventral prefrontal cortices. *The Journal of Neuroscience, 23*, 7931-7939.

O'Doherty, J., Kringelbach, M. L., Rolls, E. T., Hornak, J., & Andrews, C. (2001). Abstract reward and punishment representations in the human orbitofrontal cortex. *Nature Neuroscience, 4*, 95.

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

Ochsner, K. N., Bunge, S. A., Gross, J. J., & Gabrieli, J. D. (2002). Rethinking feelings: An fMRI study of the cognitive regulation of emotion. *Journal of Cognitive Neuroscience,* *14*, 1215-1229.

Ochsner, K. N., & Gross, J. J. (2004). Thinking makes it so: A social cognitive neuroscience approach to emotion regulation. In R. F. Baumeister & K. D. Vohs (Eds.), *Handbook of self-regulation. Research, theory and applications* (pp. 229-258). New York: The Guilford Press.

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

Ochsner, K. N., Knierim, K., Ludlow, D. H., Hanelin, J., Ramachandran, L., Glover, G., et al. (2004). Reflecting upon feelings: An fMRI study of neural systems supporting the attribution of emotion to self and other. *Journal of Cognitive Neuroscience, 16*, 1746-1772.

Ochsner, K. N., Ray, R. D., Cooper, J. C., Robertson, E. R., Chopra, S., Gabrieli, J. D., et al. (2004). For better or for worse: Neural systems supporting the cognitive down- and up-regulation of negative emotion. *NeuroImage, 23*, 483-499.

Ogawa, S., Lee, T. M., Kay, A. R., & Tank, D. W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences of the United States of America, 87*, 9868-9872.

Okuda, J., Fujii, T., Ohtake, H., Tsukiura, T., Tanji, K., Suzuki, K., et al. (2003). Thinking of the future and past: The roles of the frontal pole and the medial temporal lobes. *NeuroImage, 19*, 1369-1380.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9*, 97-113.

Olson, C. L. (1976). On choosing a test statistic in multivariate analysis of variance. *Psychological Bulletin, 83*, 579-586.

Olson, I. R., Plotzker, A., & Ezzyat, Y. (2007). The enigmatic temporal pole: A review of findings on social and emotional processing. *Brain, 130*, 1718-1731.

Öngür, 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.

Oppenheimer, S. M., Gelb, A., Girvin, J. P., & Hachinski, V. C. (1992). Cardiovascular effects of human insular cortex stimulation. *Neurology, 42*, 1727-1732.

Orth, U., Berking, M., & Burkhardt, S. (2006). Self-conscious emotions and depression: Rumination explains why shame but not guilt is maladaptive. *Personality and Social Psychology Bulletin, 32*, 1608-1619.

Osgood, C. E., Suci, G., & Tannenbaum, P. (1957). *The Measurement of Meaning*. Urbana IL: University of Illinois Press.

Ottaviani, C., Shapiro, D., Davydov, D. M., Goldstein, I. B., & Mills, P. J. (2009). The autonomic phenotype of rumination. *International Journal of Psychophysiology, 72*, 267-275.

Ottaviani, C., Shapiro, D., & Fitzgerald, L. (2011). Rumination in the laboratory: What happens when you go back to everyday life? *Psychophysiology, 48*, 453-461.

Ozgul, S., Heubeck, B., Ward, J., & Wilkinson, R. (2003). Self-discrepancies: Measurement and Relation to Various Negative Affective States. *Australian Journal of Psychology, 55*, 56-62.

Pagani, M., Mazzuero, G., Ferrari, A., Liberati, D., Cerutti, S., Vaitl, D., et al. (1991). Sympathovagal interaction during mental stress. A study using spectral analysis of heart rate variability in healthy control subjects and patients with a prior myocardial infarction. *Circulation, 83*, II43-51.

Panksepp, J. (1998). *Affective Neuroscience: The foundations of human and animal emotions*. New York: Oxford University Press.

Panksepp, J. (2005). Affective consciousness: Core emotional feelings in animals and humans. *Consciousness and Cognition, 14*, 30-80.

Paradiso, S., Robinson, R., Andreasen, N., Downhill, J., Davidson, R., Kirchner, P., et al. (1997). Emotional activation of limbic circuitry in elderly normal subjects in a PET study. *American Journal of Psychiatry, 154*, 384-389.

Pardini, D. A., Lochman, J. E., & Frick, P. J. (2003). Callous/unemotional traits and social-cognitive processes in adjudicated youths. *Journal of the American Academy for Child Adolescent Psychiatry, 42*, 364-371.

Pardo, J. V., Pardo, P. J., & Raichle, M. E. (1993). Neural correlates of self-induced dysphoria. *American Journal of Psychiatry, 150*, 713-719.

Parr, L. A., Waller, B. M., & Fugate, J. (2005). Emotional communication in primates: Implications for neuobiology. *Current Opinion in Neurobiology, 15*, 716-720.

Patterson, C. M., Kosson, D. S., & Newman, J. P. (1987). Reaction to punishment, reflectivity, and passive avoidance learning in extraverts. *Journal of Personality and Social Psychology, 52*, 565-575.

Patterson, C. M., & Newman, J. P. (1993). Reflectivity and learning from aversive events: Toward a psychological mechanism for the syndromes of disinhibition. *Psychological Review, 100*, 716-736.

Patterson, J. C., 2nd, Ungerleider, L. G., & Bandettini, P. A. (2002). Task-independent functional brain activity correlation with skin conductance changes: An fMRI study. *NeuroImage, 17*, 1797-1806.

Paus, T., Koski, L., Caramanos, Z., & Westbury, C. (1998). Regional differences in the effects of task difficulty and motor output on blood flow response in the human anterior cingulate cortex: A review of 107 PET activation studies. *Neuroreport, 9*, R37-47.

Peltz, E., Seifert, F., DeCol, R., Dörfler, A., Schwab, S., & Maihöfner, C. (2011). Functional connectivity of the human insular cortex during noxious and innocuous thermal stimulation. *NeuroImage, 54*, 1324-1335.

Penfield, W., & Faulk, M. E., Jr. (1955). The insula; further observations on its function. *Brain, 78*, 445-470.

Petrovic, P., Kalso, E., Petersson, K. M., & Ingvar, M. (2002). Placebo and opioid analgesia - imaging a shared neuronal network. *Science, 295*, 1737-1740.

Phan, K. L., Liberzon, I., Welsh, R. C., Britton, J. C., & Taylor, S. F. (2003). Habituation of rostral anterior cingulate cortex to repeated emotionally salient pictures. *Neuropsychopharmacology, 28*, 1344-1350.

Phan, K. L., Taylor, S. F., Welsh, R. C., Decker, L. R., Noll, D. C., Nichols, T. E., et al. (2003). Activation of the medial prefrontal cortex and extended amygdala by individual ratings of emotional arousal: A fMRI study. *Biological Psychiatry, 53*, 211-215.

Phan, K. L., Taylor, S. F., Welsh, R. C., Ho, S.-H., Britton, J. C., & Liberzon, I. (2004). Neural correlates of individual ratings of emotional salience: A trial-related fMRI study. *NeuroImage, 21*, 768-780.

Phan, K. L., Wager, T., 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.

Phan, K. L., Wager, T. D., Taylor, S. F., & Liberzon, I. (2004). Functional neuroimaging studies of human emotions. *CNS Spectrums, 9*, 258-266.

Phelps, E. A., & LeDoux, J. E. (2005). Contributions of the amygdala to emotion processing: From animal models to human behavior. *Neuron, 48*, 175-187.

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

Phillips, A. G., & Silvia, P. J. (2005). Self-awareness and the emotional consequences of self-discrepancies. *Personality and Social Psychology Bulletin, 31*, 703-713.

Phillips, M. L., Young, A. W., Senior, C., Brammer, M., Andrew, C., Calder, A. J., et al. (1997). A specific neural substrate for perceiving facial expressions of disgust. *Nature, 389*, 495-498.

Piaget (1965). *The moral judgment of the child*. New York: Free Press.

Piers, G., & Singer, M. B. (1953). *Shame and Guilt. A psychoanalytic and a cultural study*. Springfield, IL: Thomas.

Plant, E. A., & Devine, P. G. (1998). Internal and external motivation to respond without prejudice. *Journal of Personality and Social Psychology, 75*, 811-832.

Plant, E. A., & Devine, P. G. (2000). Expectations for interracial interactions. *Unpublished raw data*.

Plant, E. A., Devine, P. G., & Brazy, P. C. (2003). The bogus pipeline and motivations to respond without prejudice: Revisiting the fading and faking of racial prejudice. *Group Processes and Intergroup Relations, 6*, 187-200.

Plato (1949). *Timaeus* (B. Jowett, Trans.). Indianapolis, IN: Bobbs-Merrill. (Original work published 4th century B.C.)

Pole, N. (2007). The psychophysiology of posttraumatic stress disorder: A meta-analysis. *Psychological Bulletin, 133*, 725-746.

Porges, S. W. (1992). Vagal tone: A physiologic marker of stress vulnerability. *Pediatrics, 90*, 498-504.

Porges, S. W. (1995). Orienting in a defensive world: Mammalian modifications of our evolutionary heritage. A polyvagal theory. *Psychophysiology, 32*, 301-318.

Porges, S. W. (1996). Physiological regulation in high-risk infants: A model for assessment and potential intervention. *Development and Psychopathology, 8*, 43-58.

Porges, S. W. (1997). Emotion: An evolutionary by-product of the neural regulation of the autonomic nervous system. *Annals of the New York Academy of Sciences, 807*, 62-77.

Porges, S. W. (2001). The polyvagal theory: Phylogenetic substrates of a social nervous system. *International Journal of Psychophysiology, 42*, 123-146.

Porges, S. W. (2003). The Polyvagal Theory: Phylogenetic contributions to social behavior. *Physiology & Behavior, 79*, 503-513.

Porges, S. W. (2007). The polyvagal perspective. *Biological Psychology, 74*, 116-143.

Porges, S. W., & Byrne, E. A. (1992). Research methods for measurement of heart rate and respiration. *Biological Psychology, 34*, 93-130.

Porges, S. W., Doussard-Roosevelt, J. A., Portales, A. L., & Greenspan, S. I. (1996). Infant regulation of the vagal "brake" predicts child behavior problems: A psychobiological model of social behavior. *Developmental Psychobiology, 29*, 697-712.

Poskocil, A. (1977). Encounters between blacks and white liberals: The collision of stereotypes. *Social Forces, 55*, 715-727.

Price, D. D. (2000). Psychological and neural mechanisms of the affective dimension of pain. *Science, 288*, 1769.

Prohovnik, I., Skudlarski, P., Fulbright, R. K., Gore, J. C., & Wexler, B. E. (2004). Functional MRI changes before and after onset of reported emotions. *Psychiatry Research: Neuroimaging, 132*, 239-250.

Pruett, S. R., & Chan, F. (2006). The development and psychometric validation of the Disability Attitude Implicit Association Test. *Rehabilitation Psychology, 51*, 202-213.

Puce, A., Allison, T., Bentin, S., Gore, J. C., & McCarthy, G. (1998). Temporal cortex activation in humans viewing eye and mouth movements. *The Journal of Neuroscience, 18*, 2188-2199.

Quiles, Z. N., & Bybee, J. (1997). Chronic and predispositional guilt: Relations to mental health, prosocial behavior, and religiosity. *Journal of Personality Assessment, 69*, 104-126.

Rachman, S. (1993). Obsessions, responsibility and guilt. *Behaviour Research and Therapy, 31*, 149-154.

Raichle, M. E. (1994). Images of the mind: studies with modern imaging techniques. *Annual Review of Psychology, 45*, 333-356.

Raichle, M. E. (1998). Behind the scenes of functional brain imaging: A historical and physiological perspective. *Proceedings of the National Academy of Sciences of the United States of America, 95*, 765-772.

Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., & Shulman, G. L. (2001). A default mode of brain function. *Proceedings of the National Academy of Sciences of the United States of America, 98*, 676-682.

Raine, A., Buchsbaum, M. S., Stanley, J., Lottenberg, S., Abel, L., & Stoddard, J. (1994). Selective reductions in prefrontal glucose metabolism in murderers. *Biological Psychiatry, 36*, 365-373.

Rainville, P., Bechara, A., Naqvi, N., & Damasio, A. R. (2006). Basic emotions are associated with distinct patterns of cardiorespiratory activity. *International Journal of Psychophysiology, 61*, 5-18.

Rainville, P., Duncan, G. H., Price, D. D., Carrier, B., & Bushnell, M. C. (1997). Pain affect encoded in human anterior cingulate but not somatosensory cortex. *Science, 277*, 968-971.

Rauch, S. L., Shin, L. M., Segal, E., Pitman, R. K., Carson, M. A., McMullin, K., et al. (2003). Selectively reduced regional cortical volumes in post-traumatic stress disorder. *Neuroreport, 14*, 913-916.

Ravaja, N. (2004). Effects of a small talking facial image on autonomic activity: The moderating influence of dispositional BIS and BAS sensitivities and emotions. *Biological Psychology, 65*, 163-183.

Redelmeier, D. A., & Singh, S. M. (2001). Survival in Academy Award-winning actors and actresses. *Annals of Internal Medicine, 134*, 955-962.

Reiman, E. M., Lane, R. D., Ahern, G. L., Schwartz, G. E., Davidson, R. J., Friston, K. J., et al. (1997). Neuroanatomical correlates of externally and internally generated human emotion. *American Journal of Psychiatry, 154*, 918-925.

Remijnse, P. L., Nielen, M. M. A., Uylings, H. B. M., & Veltman, D. J. (2005). Neural correlates of a reversal learning task with an affectively neutral baseline: An event-related fMRI study. *NeuroImage, 26*, 609-618.

Riese, H., Groot, P. F. C., Van den Berg, M., Kupper, N. H. M., Magnee, E. H. B., Rohaan, E. J., et al. (2003). Large-scale ensemble averaging of ambulatory impedance cardiograms. *Behavior Research Methods, Instruments, and Computers, 35*, 467-477.

Ritz, T., Alatupa, S., Thons, M., & Dahme, B. (2002). Effects of affective picture viewing and imagery on respiratory resistance in nonasthmatic individuals. *Psychophysiology, 39*, 86-94.

Roberts, N. A., Levenson, R. W., & Gross, J. J. (2008). Cardiovascular costs of emotion suppression cross ethnic lines. *International Journal of Psychophysiology, 70*, 82-87.

Roberts, N. A., Tsai, J. L., & Coan, J. A. (2007). Emotion elicitation using dyadic interaction tasks. In J. A. Coan & J. J. B. Allen (Eds.), *Handbook of emotion elicitation and assessment* (pp. 106-123). New York: Cambridge University Press.

Robins, R. W., Noftle, E. E., & Tracy, J. L. (2007). Assessing self-conscious emotions: A review of self-report and nonverbal measures. In J. L. Tracy, R. W. Robins & J. P. Tangney (Eds.), *The self-conscious emotions: Theory and research*. New York: Guilford.

Robins, R. W., Trzesniewski, K. H., Tracy, J. L., Gosling, S. D., & Potter, J. (2002). Global self-esteem across the life span. *Psychology and Aging, 17*, 423-434.

Rojahn, J., Komelasky, K., & Man, M. (Writer) (2008). Implicit Attitudes and Explicit Ratings of Romantic Attraction of College Students Toward Opposite-Sex Peers with Physical Disabilities, *Journal of Developmental and Physical Disabilities*: Springer Science & Business Media B.V.

Rolls, E. T. (1999). *The brain and emotion*. New York: Oxford University Press.

Rolls, E. T. (2000). The orbitofrontal cortex and reward. *Cerebral Cortex, 10*, 284-294.

Rotshtein, P., Henson, R. N., Treves, A., Driver, J., & Dolan, R. J. (2005). Morphing Marilyn into Maggie dissociates physical and identity face representations in the brain. *Nature Neuroscience, 8*, 107-113.

Rottenberg, J. (2007). Cardiac vagal control in depression: A critical analysis. *Biological Psychology, 74*, 200-211.

Rottenberg, J., Kasch, K. L., Gross, J. J., & Gotlib, I. H. (2002). Sadness and amusement reactivity differentially predict concurrent and prospective functioning in major depressive disorder. *Emotion*, *2*, 135-146.

Rottenberg, J., Ray, R. D., & Gross, J. J. (2007). Emotion elicitation using films. In J. A. Coan & J. J. B. Allen (Eds.), *Handbook of emotion elicitation and assessment* (pp. 9-28). New York: Oxford University Press.

Rozin, P., Haidt, J., & McCauley, C. R. (2000). Disgust. In M. Lewis & J. M. Haviland-Jones (Eds.), *Handbook of emotions* (2nd ed., pp. 637-653). New York: Guilford Press.

Ryaff, C. D., & Singer, B. H. (2003). The role of emotion on pathways to positive health. In R. J. Davidson, K. R. Scherer & H. H. Goldsmith (Eds.), *Handbook of affective sciences* (pp. 1083-1104). New York: Oxford University Press.

Ryan, R. M., & Connell, J. P. (1989). Perceived locus of causality and internalization: Examining reasons for acting in two domains. *Journal of Personality and Social Psychology, 57*, 749-761.

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology, 25*, 54-67.

Sabatinelli, D., Lang, P. J., Bradley, M. M., Costa, V. D., & Keil, A. (2009). The timing of emotional discrimination in human amygdala and ventral visual cortex. *The Journal of Neuroscience, 29*, 14864-14868.

Sachs-Ericsson, N., & Ciarlo, J. A. (2000). Gender, social roles, and mental health: An epidemiological perspective. *Sex Roles, 43*, 605-628.

Salomons, T. V., Johnstone, T., Backonja, M. M., & Davidson, R. J. (2004). Perceived controllability modulates the neural response to pain. *The Journal of Neuroscience, 24*, 7199-7203.

Sánchez-Navarro, J. P., Martínez-Selva, J. M., & Román, F. (2006). Uncovering the relationship between defence and orienting in emotion: Cardiac reactivity to unpleasant pictures. *International Journal of Psychophysiology, 61*, 34-46.

Sander, M., Macefield, V. G., & Henderson, L. A. (2010). Cortical and brain stem changes in neural activity during static handgrip and postexercise ischemia in humans. *Journal of Applied Physiology, 108*, 1691-1700.

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

Saper, C. B. (2002). The central autonomic nervous system: Conscious visceral perception and autonomic pattern generation. *Annual Review of Neuroscience, 25*, 433.

Satpute, A. B., & Lieberman, M. D. (2006). Integrating automatic and controlled processes into neurocognitive models of social cognition. *Brain Research, 1079*, 86-97.

Saver, J. L., & Damasio, A. R. (1991). Preserved access and processing of social knowledge in a patient with acquired sociopathy due to ventromedial frontal damage. *Neuropsychologia, 29*, 1241-1249.

Saxe, R. (2006). Four brain regions for one Theory of Mind. In J. T. Cacioppo, P. S. Visser & C. L. Pickett (Eds.), *Social neuroscience: People thinking about thinking people* (pp. 83-101). Cambridge: MIT Press.

Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people: The role of the temporo-parietal junction in "theory of mind". *NeuroImage, 19*, 1835-1842.

Schachter, S., & Singer, J. E. (1962). Cognitive, social, and physiological determinants of emotional state. *Psychological Review, 69*, 379-399.

Schmader, T., & Lickel, B. (2006). The approach and avoidance function of guilt and shame emotions: Comparing reactions to self-caused and other-caused wrongdoing. *Motivation and Emotion, 30*, 43-56.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002a). *E-Prime reference guide*. Pittsburgh: Psychology Sofware Tools Inc.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002b). *E-Prime user's guide*. Pittsburgh: Psychology Software Tools Inc.

Schwartz, G. E., Weinberger, D. A., & Singer, J. A. (1981). Cardiovascular differentiation of happiness, sadness, anger, and fear following imagery and exercise. *Psychosomatic Medicine, 43*, 343-364.

Seeley, W. W., Menon, V., Schatzberg, A. F., Keller, J., Glover, G. H., Kenna, H., et al. (2007). Dissociable intrinsic connectivity networks for salience processing and executive control. *The Journal of Neuroscience, 27*, 2349-2356.

Seymour, B., Singer, T., & Dolan, R. (2007). The neurobiology of punishment. *Nature Reviews Neuroscience, 8*, 300-312.

Shafritz, K. M., Collins, S. H., & Blumberg, H. P. (2006). The interaction of emotional and cognitive neural systems in emotionally guided response inhibition. *NeuroImage, 31*, 468-475.

Sheikh, S., & Janoff-Bulman, R. (2010). The "shoulds" and "should nots" of moral emotions: A self-regulatory perspective on shame and guilt. *Personality and Social Psychology Bulletin, 36*, 213-224.

Sherman, S. J., & Gorkin, L. (1980). Attitude bolstering when behavior is inconsistent with central attitudes. *Journal of Experimental Social Psychology, 16*, 388-403.

Sherwood, A., Allen, M. T., Fahrenberg, J., Kelsey, R. M., Lovallo, W. R., & Van Doornen, L. J. P. (1990). Methodological guidelines for impedance cardiography. *Psychophysiology, 27,* 1-23.

Shields, S. A. (1991). Gender in the psychology of emotion: A selective research review. In K. T. Strongman (Ed.), *International review of studies on emotion* (pp. 227-245). New York: Wiley.

Shin, L. M., Dougherty, D. D., Orr, S. P., Pitman, R. K., Lasko, M., Macklin, M. L., et al. (2000). Activation of anterior paralimbic structures during guilt-related script-driven imagery. *Biological Psychiatry, 48*, 43-50.

Shin, L. M., Orr, S. P., Margaret, A. C., Rauch, S. L., Macklin, M. L., Lasko, N. B., et al. (2004). Regional cerebral blood flow in the amygdala and medial prefrontal cortex during traumatic imagery in male ad female Vietnam veterans with PTSD. *Archives of General Psychiatry, 61*, 168-176.

Siegle, G. J., Steinhauer, S. R., Thase, M. E., Stenger, V. A., & Carter, C. S. (2002). Can't shake that feeling: Event-related fMRI assessment of sustained amygdala activity in response to emotional information in depressed individuals. *Biological Psychiatry, 51*, 693-707.

Simpson, J. R., Jr., Snyder, A. Z., Gusnard, D. A., & Raichle, M. E. (2001). Emotion-induced changes in human medial prefrontal cortex: I. During cognitive task performance. *Proceedings of the National Academy of Sciences of the United States of America, 98*, 683-687.

Singer, T., Seymour, B., O'Doherty, J., Kaube, H., Dolan, R. J., & Frith, C. D. (2004). Empathy for pain involves the affective but not sensory components of pain. *Science, 303*, 1157-1162.

Singer, W. (2001). Consciousness and the binding problem. *Annals of the New York Academy of Sciences, 929*, 123-146.

Sleight, P., La Rovere, M. T., Mortara, A., Pinna, G., Maestri, R., Leuzzi, S., et al. (1995). Physiology and pathophysiology of heart rate and blood pressure variability in humans: Is power spectral analysis largely an index of baroreflex gain? *Clinical Science (London, England: 1979), 88*, 103-109.

Smillie, L. D., Pickering, A. D., & Jackson, C. J. (2006). The new reinforcement sensitivity theory: Implications for personality measurement. *Personality and Social Psychology Review 10*, 320-335.

Smith, C. A., & Pope, L. K. (1992). Appraisal and emotion: The interactional contributions of dispositional and situational factors. In M. S. Clark (Ed.), *Review of personality and social psychology: Emotion and social behaviour* (Vol. 14, pp. 32-62). Newbury Park, CA: Sage.

Smith, R. H., Webster, J. M., Parrott, W. G., & Eyre, H. L. (2002). The role of public exposure in moral and nonmoral shame and guilt. *Journal of Personality and Social Psychology, 83*, 138-159.

Solms, M., & Turnbull, O. (2002). *The brain and the inner world. An introduction to the neuroscience of subjective experience.* New York: Other Press.

Somerville, L. H., Kim, H., Johnstone, T., Alexander, A. L., & Whalen, P. J. (2004). Human amygdala responses during presentation of happy and neutral faces: Correlations with state anxiety. *Biological Psychiatry, 55*, 897-903.

Sommer, M., Rothmayr, C., Dohnel, K., Meinhardt, J., Schwerdtner, J., Sodian, B., et al. (2010). How should I decide? The neural correlates of everyday moral reasoning. *Neuropsychologia, 48*, 2018-2026.

Somsen, R. J., Van der Molen, M. W., Jennings, J. R., & van Beek, B. (2000). Wisconsin Card Sorting in adolescents: Analysis of performance, response times and heart rate. *Acta Psychologica 104*, 227-257.

Sprengelmeyer, R., Rausch, M., Eysel, U. T., & Przuntek, H. (1998). Neural structures associated with recognition of facial expressions of basic emotions. *Proceedings of the Royal Society B: Biological Sciences*, *265*, 1927-1931.

Steele, J. D., & Lawrie, S. M. (2004). Segregation of cognitive and emotional function in the prefrontal cortex: A stereotactic meta-analysis. *NeuroImage, 21*, 868-875.

Stellar, E. (1994). The physiology of motivation. 1954. *Psychological Review, 101*, 301-311.

Stemmler, G. (1989). The autonomic differentiation of emotions revisited: Convergent and discriminant validation. *Psychophysiology, 26*, 617-632.

Stemmler, G., Aue, T., & Wacker, J. (2007). Anger and fear: Separable effects of emotion and motivational direction on somatovisceral responses. *International Journal of Psychophysiology, 66*, 141-153.

Stemmler, G., Heldman, M., Pauls, C. A., & Scherer, T. (2001). Constraints for emotion specificity in fear and anger: The context counts. *Psychophysiology, 38*, 275-291.

Steptoe, A., & Brydon, L. (2009). Emotional triggering of cardiac events. *Neuroscience and Biobehavioural Reviews, 33*, 63-70.

Steptoe, A., Fieldman, G., Evans, O., & Perry, L. (1996). Cardiovascular risk and responsivity to mental stress: The influence of age, gender and risk factors. *Journal of Cardiovascular Risk, 3*, 83-93.

Strahan, R., & Gerbasi, K. C. (1972). Short, homogenous versions of the Marlowe-Crowne

Social Desirability Scale. *Journal of Clinical Psychology, 28*, 191-193.

Strauman, T. J. (1992). Self-guides, autobiographical memory, and anxiety and dysphoria: Toward a cognitive model of vulnerability to emotional distress. *Journal of Abnormal Psychology, 101*, 87-95.

Sturm, V. E., Ascher, E. A., Miller, B. L., & Levenson, R. W. (2008). Diminished self-conscious emotional responding in frontotemporal lobar degeneration patients. *Emotion, 8*, 861-869.

Sturm, V. E., Rosen, H. J., Allison, S., Miller, B. L., & Levenson, R. W. (2006). Self-conscious emotion deficits in frontotemporal lobar degeneration. *Brain, 129*, 2508-2516.

Sullivan, M. W., Bennett, D. S., & Lewis, M. (2003). Darwin's view: Self-evaluative emotions as context-specific emotions. *Annals of the New York Academy of Sciences, 1000*, 304-308.

Takahashi, H., Matsuura, M., Koeda, M., Yahata, N., Suhara, T., Kato, M., et al. (2008). Brain activations during judgments of positive self-conscious emotion and positive basic emotion: Pride and joy. *Cerebral Cortex, 18,* 898-903.

Takahashi, H., Matsuura, M., Yahata, N., Koeda, M., Suhara, T., & Okubo, Y. (2006). Men and women show distinct brain activations during imagery of sexual and emotional infidelity. *NeuroImage, 32,* 1299-1307.

Takahashi, H., Yahata, N., Koeda, M., Matsuda, T., Asai, K., & Okubo, Y. (2004). Brain activation associated with evaluative processes of guilt and embarrassment: An fMRI study. *NeuroImage, 23*, 967-974.

Tangney, J. P. (1990). Assessing individual differences in proneness to shame and guilt: Development of the Self-Conscious Affect and Attribution Inventory. *Journal of Personality and Social Psychology, 59*, 102-111.

Tangney, J. P. (1991). Moral affect: The good, the bad, and the ugly. *Journal of Personality and Social Psychology, 61*, 598-607.

Tangney, J. P. (1992). Situational determinants of shame and guilt in young adulthood. *Personality and Social Psychology Bulletin, 18*, 199-206.

Tangney, J. P. (1995). Recent advances in the empirical study of shame and guilt. *American Behavioral Scientist, 38*, 1132-1145.

Tangney, J. P. (1996). Conceptual and methodological issues in the assessment of shame and guilt. *Behavioural Research Therapy, 34*, 741-754.

Tangney, J. P., & Dearing, R. L. (2002). *Shame and guilt*. New York: The Guilford Press.

Tangney, J. P., Dearing, R. L., Wagner, P. E., & Gramzow, R. (2000). *The Test of Self-Conscious Affect-3 (TOSCA-3)*. Fairfax, VA: George Mason University.

Tangney, J. P., Miller, R. S., Flicker, L., & Barlow, D. H. (1996). Are shame, guilt, and embarrassment distinct emotions? *Journal of Personality and Social Psychology, 70,* 1256-1269.

Tangney, J. P., Niedenthal, P. M., Covert, M. V., & Barlow, D. H. (1998). Are shame and guilt related to distinct self-discrepancies? A test of Higgins's (1987) hypotheses. *Journal of Personality and Social Psychology, 75,* 256-268.

Tangney, J. P., Struewig, J., & Mashek, D. J. (2007). Moral emotions and moral behavior. *Annual Review of Psychology, 58,* 345-372.

Tangney, J. P., Wagner, P., Fletcher, C., & Gramzow, R. (1992). Shamed into anger? The relation of shame and guilt to anger and self-reported aggression. *Journal of Personality and Social Psychology, 62*, 669-675.

Tangney, J. P., Wagner, P., & Gramzow, R. (1992). Proneness to shame, proneness to guilt, and psychopathology. *Journal of Abnormal Psychology, 101,* 469-478.

Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology (1996). Heart rate variability: Standards of measurement, physiological interpretation, and clinical use. *European Heart Journal*, *17*, 354-381.

Taylor, K. S., Seminowicz, D. A., & Davis, K. D. (2009). Two systems of resting state connectivity between the insula and cingulate cortex. *Human Brain Mapping, 30,* 2731-2745.

Taylor, S. E. (1991). Asymmetrical effects of positive and negative events: The mobilization-minimization hypothesis. *Psychological Bulletin, 110*, 67-85.

Teroni, F., & Deonna, J. A. (2008). Differentiating shame from guilt. *Consciousness and Cognition, 17,* 725-740.

Thayer, J. F., & Lane, R. D. (2000). A model of neurovisceral integration in emotion regulation and dysregulation. *Journal of Affective Disorders, 61*, 201-216.

Thayer, J. F., & Lane, R. D. (2007). The role of vagal function in the risk for cardiovascular disease and mortality. *Biological Psychology*, *74*, 224-242.

Thayer, J. F., & Lane, R. D. (2009). Claude Bernard and the heart-brain connection: Further elaboration of a model of neurovisceral integration. *Neuroscience and Biobehavioral Reviews, 33*, 81-88.

Thompson, E. R. (2007). Positive and Negative Affect Schedule (PANAS): Development and validation of an internationally reliable short-form of the Positive and Negative Affect Schedule (PANAS). *Journal of Cross-Cultural Psychology, 38*, 227-242.

Thompson, R. J., & Berenbaum, H. (2006). Shame reactions to everyday dilemmas are associated with depressive disorder. *Cognitive Therapy and Research, 30,* 415-425.

Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., et al. (2009). The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Research*, *168*, 242-249.

Tracy, J. L., & Robins, R. W. (2004a). Puting the self into self-conscious emotions: A theoretical model. *Psychological Inquiry*, *15*, 103-125.

Tracy, J. L., & Robins, R. W. (2004b). Show your pride: Evidence for a discrete emotion expression. *Psychological Science, 15*, 194-197.

Tracy, J. L., & Robins, R. W. (2007a). Emerging insights into the nature and function of pride. *Association for Psychological Science, 16*, 147-150.

Tracy, J. L., & Robins, R. W. (2007b). The psychological structure of pride: A tale of two facets. *Journal of Personality and Social Psychology, 92*, 506-525.

Tracy, J. L., & Robins, R. W. (2008). The nonverbal expression of pride: Evidence for cross-cultural recognition. *Journal of Personality and Social Psychology, 94*, 516-530.

Tranel, D. (1983). The effects of monetary incentive and frustrative nonreward on heart rate and electrodermal activity. *Psychophysiology*, *20*, 652-657.

Tranel, D. (1994). "Acquired sociopathy": The development of sociopathic behavior following focal brain damage. *Progress in Experimental Personality and Psychopathology Research*, 285-311.

Tugade, M. M., Fredrickson, B. L., & Barrett, L. F. (2004). Psychological resilience and positive emotional granularity: Examining the benefits of positive emotions on coping and health. *Journal of Personality, 72*, 1161-1190.

Turner, T. J., & Ortony, A. (1992). Basic emotions: Can conflicting criteria converge? *Psychological Review, 99*, 566-571.

Valdesolo, P., & DeSteno, D. (2006). Manipulations of emotional context shape moral judgment. *Psychological Science 17*, 476-477.

van der Veen, F. M., van der Molen, M. W., Crone, E. A., & Jennings, J. R. (2004). Phasic heart rate responses to performance feedback in a time production task: Effects of information versus valence. *Biological Psychology, 65*, 147-161.

van Lang, N., Tulen, J., Kallen, V., Rosbergen, B., Dieleman, G., & Ferdinand, R. (2007). Autonomic reactivity in clinically referred children attention-deficit/hyperactivity disorder versus anxiety disorder. *European Child and Adolescent Psychiatry, 16*, 71-78.

van Veen, V., & Carter, C. S. (2002). The timing of action-monitoring processes in the anterior cingulate cortex. *Journal of Cognitive Neuroscience, 14*, 593.

van Veen, V., & Carter, C. S. (2006). Error detection, correction, and prevention in the brain: A brief review of data and theories. *Clinical EEG and Neuroscience, 37*, 330-335.

van Veen, V., Cohen, J. D., Botvinick, M. M., Stenger, V. A., & Carter, C. S. (2001). Anterior cingulate cortex, conflict monitoring, and levels of processing. *NeuroImage, 14*, 1302-1308.

Verduyn, P., Delvaux, E., Van Coillie, H., Tuerlinckx, F., & Van Mechelen, I. (2009). Predicting the duration of emotional experience: Two experience sampling studies. *Emotion, 9*, 83-91.

Viinikainen, M., Jääskeläinen, I. P., Alexandrov, Y., Balk, M. H., Autti, T., & Sams, M. (2010). Nonlinear relationship between emotional valence and brain activity: Evidence of separate negative and positive valence dimensions. *Human Brain Mapping, 31*, 1030-1040.

Vogt, B. A., Finch, D. M., & Olson, C. R. (1992). Functional heterogeneity in cingulate cortex: The anterior executive and posterior evaluative regions. *Cerebral Cortex, 2*, 435-443.

Völlm, B. A., Taylor, A. N., Richardson, P., Corcoran, R., Stirling, J., McKie, S., et al. (2006). Neuronal correlates of theory of mind and empathy: A functional magnetic resonance imaging study in a nonverbal task. *NeuroImage, 29*, 90-98.

Vrijkotte, T. G., van Doornen, L. J., & de Geus, E. J. (2004). Overcommitment to work is associated with changes in cardiac sympathetic regulation. *Psychosomatic Medicine, 66*, 656-663.

Vuilleumier, P., & Pourtois, G. (2007). Distributed and interactive brain mechanisms during emotion face perception: Evidence from functional neuroimaging. *Neuropsychologia, 45*, 174-194.

Vul, E., Harris, C., Winkielman, P., & Pashler, H. (2009). Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition. *Perspectives on Psychological Science, 4*, 274-290.

Wager, T. D., Phan, K. L., Liberzon, I., & Taylor, S. F. (2003). Valence, gender, and lateralization of functional brain anatomy in emotion: A meta-analysis of findings from neuroimaging. *NeuroImage, 19*, 513-531.

Wager, T. D., Rilling, J. K., Smith, E. E., Sokolik, A., Casey, K. L., Davidson, R. J., et al. (2004). Placebo-induced changes in fMRI in the anticipation and experience of pain. *Science, 303*, 1162-1167.

Waldstein, S. R., Kop, W. J., Schmidt, L. A., Haufler, A. J., Krantz, D. S., & Fox, N. A. (2000). Frontal electrocortical and cardiovascular reactivity during happiness and anger. *Biological Psychology, 55*, 3-23.

Wall, P. D. (1999). *Pain: The science of suffering*. London: Weidenfeld & Nicolson.

Wallin, B. G., Esler, M., Dorward, P., Eisenhofer, G., Ferrier, C., Westerman, R., et al. (1992). Simultaneous measurements of cardiac noradrenaline spillover and sympathetic outflow to skeletal muscle in humans. *The Journal of Physiology, 453*, 45-58.

Ward, C. L., Flisher, A. J., Zissis, C., Muller, M., & Lombard, C. (2001). Exposure to violence and its relationship to psychopathology in adolescents. *Injury Prevention, 7*, 297-301.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54*, 1063-1070.

Weber, E. J. M., Van Der Molen, M. W., & Molenaar, P. C. M. (1994). Heart rate and sustained attention during childhood: Age changes in anticipatory heart rate, primary bradycardia, and respiratory sinus arrhythmia. *Psychophysiology, 31*, 164-174.

Webster, J. M., Duvall, J., Gaines, L. M., & Smith, R. H. (2003). The roles of praise and social comparison information in the experience of pride. *The Journal of Social Psychology, 143*, 209-232.

Werner, E. (2000). Protective factors and individual resilience. In R. Meisells & J. Shonkoff (Eds.), *Handbook of early intervention* (pp. 115-132). Cambridge: Cambridge University Press.

Whalen, P. J., Bush, G., McNally, R. J., Wilhelm, S., McInerney, S. C., Jenike, M. A., et al. (1998). The emotional counting Stroop paradigm: A functional magnetic resonance imaging probe of the anterior cingulate affective division. *Biological Psychiatry, 44*, 1219-1228.

Whisman, M. A., Perez, J. E., & Ramel, W. (2000). Factor structure of the Beck Depression Inventory second edition (BDI-II) in a student sample. *Journal of Clinical Psychology, 56*, 545-551.

Wicker, B., Keysers, C., Plailly, J., Royet, J. P., Gallese, V., & Rizzolatti, G. (2003). Both of us disgusted in my insula: The common neural basis of seeing and feeling disgust. *Neuron, 40*, 655-664.

Williams, D. R., Gonzalez, H. M., Williams, S., Mohammed, S. A., Moomal, H., & Stein, D. J. (2008). Perceived discrimination, race and health in South Africa. *Social Science and Medicine, 67*, 441-452.

Williams, L. A., & DeSteno, D. (2008). Pride and perseverance: The motivational role of pride. *Journal of Personality and Social Psychology, 94*, 1007-1017.

Williams, L. M., Das, P., Liddell, B., Olivieri, G., Peduto, A., Brammer, M. J., et al. (2005). BOLD, sweat and fears: fMRI and skin conductance distinguish facial fear signals. *Neuroreport, 16*, 49-52.

Williamson, J. W., Fadel, P. J., & Mitchell, J. H. (2006). New insights into central cardiovascular control during exercise in humans: A central command update. *Experimental Physiology, 91*, 51-58.

Williamson, J. W., McColl, R., & Mathews, D. (2003). Evidence for central command activation of the human insular cortex during exercise. *Journal of Applied Physiology, 94*, 1726-1734.

Williamson, J. W., Nobrega, A. C., McColl, R., Mathews, D., Winchester, P., Friberg, L., et al. (1997). Activation of the insular cortex during dynamic exercise in humans. *The Journal of Physiology, 503*, 277-283.

Wilson, T. D., & Gilbert, D. T. (2005). Affective forecasting: Knowing what to want. *Current Directions in Psychological Science 14*, 131-134.

Wilson, T. D., Lindsey, S., & Schooler, T. Y. (2000). A model of dual attitudes. *Psychological Review, 107*, 101-126.

Wingrove, J., & Bond, A. J. (1998). Angry reactions to failure on a cooperative computer game: The effect of trait hostility, behavioural inhibition, and behavioural activation. *Aggressive Behavior, 24*, 27-36.

Wong, S. W., Massé, N., Kimmerly, D. S., Menon, R. S., & Shoemaker, J. K. (2007). Ventral medial prefrontal cortex and cardiovagal control in conscious humans. *NeuroImage, 35*, 698-708.

Wood, J. N., & Grafman, J. (2003). Human prefrontal cortex: Processing and representational perspectives. *Nature Reviews Neuroscience, 4*, 139.

Wright, P., He, G., Shapira, N. A., Goodman, W. K., & Liu, Y. (2004). Disgust and the insula: fMRI responses to pictures of mutilation and contamination. *Neuroreport, 15*, 2347-2351.

Wu, G., Arzeno, N., Shen, L., Tang, D., Zheng, D., Zhao, N., et al. (2009). Chaotic signatures of heart rate variability and its power spectrum in health, aging and heart failure. *PLoS One, 4*, e4323.

Yeragani, V. K., Pohl, R., Berger, R., Balon, R., Ramesh, C., Glitz, D., et al. (1993). Decreased heart rate variability in panic disorder patients: A study of power-spectral analysis of heart rate. *Psychiatry Research, 46*, 89-103.

Young, L., & Saxe, R. (2008). The neural basis of belief encoding and integration in moral judgment. *NeuroImage, 40*, 1912-1920.

Zahn, R., de Oliveira-Souza, R., Bramati, I., Garrido, G., & Moll, J. (2009). Subgenual cingulate activity reflects individual differences in empathic concern. *Neuroscience Letters, 457*, 107-110.

Zahn, R., Moll, J., Krueger, F., Huey, E. D., Garrido, G., & Grafman, J. (2007). Social concepts are represented in the superior anterior temporal cortex. *Proceedings of the National*

Academy of Sciences of the United States of America, 104, 6430-6435.

Zahn, R., Moll, J., Paiva, M., Garrido, G., Krueger, F., Huey, E. D., et al. (2009). The neural basis of human social values: Evidence from functional MRI. *Cerebral Cortex, 19*, 276-283.

Zahn, T. P., Grafman, J., & Tranel, D. (1999). Frontal lobe lesions and electrodermal activity: Effects of significance. *Neuropsychologia*, *37*, 1227-1241.

Zajonc, R. B., & McIntosh, D. N. (1992). Emotions research: Some promising questions and some questionable promises. *Psychological Science, 3*, 70-74.

Zald, D. H. (2003). The human amygdala and the emotional evaluation of sensory stimuli. *Brain Research Reviews, 41*, 88-123.

Zhang, Z., & Oppenheimer, S. M. (1997). Characterization, distribution and lateralization of baroreceptor-related neurons in the rat insular cortex. *Brain Research, 760*, 243-250.

Zhang, Z. H., Rashba, S., & Oppenheimer, S. M. (1998). Insular cortex lesions alter baroreceptor sensitivity in the urethane-anesthetized rat. *Brain Research, 813*, 73-81.

APPENDICES

Appendix A: Demographic and Health Questionnaire

Participant ID: ______________________________ **Date:** ________________

1. What age are you now? __________________________ 2. Date of birth? ______________
3. Population group (circle): African Black Indian Coloured White
4. What is your first/home language? _________________
5. Which hand do you write with? __________________

Did you always use this hand since you were a child? Y N

6. Number of years of formal education: ___________________
7. What is your current job? ___________________________
8. Medical History

Have you ever been hospitalised for a neurological problem, such as head injury, epilepsy, stroke or brain tumour? Y N
What for? _______________________________

9. Have you ever been treated for problems with alcohol or drugs? _____________________
10. Psychiatric/mental health

Have you ever been prescribed drugs for a psychiatric or other mental health problem, such as anxiety, depression, or schizophrenia? Y N
When was this? When did you stop taking the drugs (if appropriate)? ___________________

11. Menstrual cycle (if applicable)

Days since last period: ______________
Length of cycle: ___________________
Do you use oral contraceptives? Y N

12. Learning disabilities

Did you have problems at school for which you received special help, such as reading problems, for instance? Y N
Please specify: ____________________________

13. Are you currently on any medication? Y N

If so, what for? ____________________________

Appendix B: Study 1 Consent Form

UNIVERSITY OF CAPE TOWN
DEPARTMENT OF PSYCHOLOGY

Informed Consent to Participate in Research and Authorization for Collection, Use, and Disclosure of Cognitive Performance and Other Personal Data

You are being asked to take part in a research study. This form provides you with information about the study and seeks your authorization for the collection, use and disclosure of your cognitive performance data, as well as other information necessary for the study. The Principal Investigator (the person in charge of this research) or a representative of the Principal Investigator will also describe this study to you and answer all of your questions. Your participation is entirely voluntary. Before you decide whether or not to take part, read the information below and ask questions about anything you do not understand. By participating in this study you will not be penalized or lose any benefits to which you would otherwise be entitled.

1. Name of Participant

____

2. Title of Research Study

Cognitive and Physiological Processes Underlying Performance on Three Psychological Tests.

3. Contact Details

Investigator: Melike Fourie

Department of Psychology
University of Cape Town
Office: (021) 650-3415
Mobile: 083 781 6021

4. Source of Funding or Other Material Support

National Research Foundation (NRF)
Supervisor Private Research Funds

5. **What is the purpose of this research study?**

The main purpose of this research essentially involves understanding how non-specific physiological arousal relates to cognitive performance in some specific neuropsychological tests. Additionally, we'd like to correlate this performance with some personality traits like general mood and attitude.

6. **What will be done if you take part in this research study?**

You will essentially be required to participate in two legs of the same study: i.e. the cognitive and affective parts, which will take place on one occasion. In the first session you will be asked to complete some questionnaires that should not take longer than about 30min. These measure certain aspects of your personality and attitude.
In the second session you will be administered a series of relatively simple cognitive tasks that are presented on a computer screen while we monitor your physiological arousal. This should take no more than 1hour. The tasks measure certain aspects of your visual, attentional, motor, and memory functioning, as well as your general cognitive functioning.
To record the physiological data, a series of sensors/electrodes will be attached to your upper body and hands to measure things like heart rate and breathing.
After the second experimental session is over, you will be informed in detail about the design of the study and the research questions we hope to answer. You will also have the opportunity to ask questions and thus learn more about psychological research.
If you have any questions now or at any time during the study, you may contact the Investigator listed in #3 of this form.

7. **If you choose to participate in this study, how long will you be expected to participate in the research?**

The experiment consists of two sessions, which together should not last longer than one and a half (1½) hours. If at any time during the experimental sessions you find any of the procedures uncomfortable, you are free to discontinue your participation without penalty.

8. **How many people are expected to participate in the research?**

About 60.

9. **What are the possible discomforts and risks?**

There are no known risks associated with participation in this study. Attaching the physiological electrodes might be experienced as uncomfortable and intrusive by some people, but this will be done by a trained female assistant and all care will be taken to be as unobtrusive as possible. You also may experience slight fatigue during the experimental sessions. However, if you become tired you will be allowed to take breaks between blocks of tests whenever you want to.
If you wish to discuss the information above or any discomforts you may experience, you may ask questions now or call the Investigator listed on the front page of this form.

10a. What are the possible benefits to you?

You may or may not personally benefit from participating in this study. Participation in this study may, however, improve your mental test performance due to training and practice.

10b. What are the possible benefits to others?

The information from this study may help improve our understanding of mental functions and cognitive processing. Additionally, this research will allow us to gather information about the performance of healthy adults on the administered tests.

11. If you choose to take part in this research study, will it cost you anything?

Participating in this study will not cost you anything.

12. Will you receive compensation for taking part in this research study?

Yes, you will receive financial compensation of R30 for taking part in this study.

13a. Can you withdraw from this research study?

You are free to withdraw your consent and to stop participating in this research study at any time. If you do withdraw your consent, there will be no penalty.

If you have any questions regarding your rights as a research subject, you may phone the Psychology Department offices at (021) 650-3430.

13b. If you withdraw, can information about you still be used and/or collected?

Information already collected may be used.

14. Once personal and performance information is collected, how will it be kept secret (confidential) in order to protect your privacy?

Information collected will be stored in locked filing cabinets or in computers with security passwords. Only certain people have the right to review these research records. These people include the researchers for this study and certain University of Cape Town officials. Your research records will not be released without your permission unless required by law or a court order.

15. What information about you may be collected, used and shared with others?

This information gathered from you will be demographic and health history information, records of your performance on certain cognitive tests, physiological data, and records of your responses to attitude questionnaires. If you agree to be in this research study, it is possible that some of the information collected might be copied into a “limited data set” to be used for other research purposes. If so, the limited data set may only include information that does not directly identify you. For example, the limited data set cannot include your name, address, telephone number, ID number, or any other photographs, numbers, codes, or so forth that link you to the information in the limited data set.

16. How will the researcher(s) benefit from your being in the study?

In general, presenting research results helps the career of a scientist. Therefore, the Principal Investigator and others attached to this research project may benefit if the results of this study are presented at scientific meetings or in scientific journals.

17. Signatures

As a representative of this study, I have explained to the participant the purpose, the procedures, the possible benefits, and the risks of this research study; and how the participant's performance and other data will be collected, used, and shared with others:

____ ____________________

Signature of Person Obtaining Consent and Authorization Date

You have been informed about this study's purpose, procedures, possible benefits, and risks; and how your performance and other data will be collected, used and shared with others. You have received a copy of this form. You have been given the opportunity to ask questions before you sign, and you have been told that you can ask other questions at any time.

You voluntarily agree to participate in this study. You hereby authorize the collection, use and sharing of your performance and other data. By signing this form, you are not waiving any of your legal rights.

____ ____________________

Signature of Person Consenting and Authorizing Date

Please indicate below if you would like to be notified of future research projects conducted by our research group:

______________ (initial) Yes, I would like to be added to your research participation pool and be notified of research projects in which I might participate in the future.

Method of contact:
Phone number: __________________________
E-mail address: __________________________
Mailing address: ____
____

Appendix C: Study 1 Cover Story and Dialogue

The section below presents the cover story that was given to all participants in Study 1 and goes on to describe the details of the experimental emotion manipulations. Whereas the indented texts represent the verbatim dialogue that was employed, the texts inbetween describe the accompanying actions and proceedings.

The following individuals were involved as 'actors':

Experimenter: Melike (dressed professionally)
Confederates:
Research assistant: Diane (she wore a white laboratory coat)
Supervisor: Haidee (she also dressed professionally, with no bright colours)

Cover Story

Experimenter: Greetings and welcome. I will now briefly explain to you the nature of the study you will participate in and then leave you in the trusted care of my assistant, Diane. This afternoon you will participate in two legs of the same study. In the first part you will be completing some questionnaires and forms about yourself, and in the second part we are interested to see how physiological measures, like heart rate and breathing, vary with cognitive function. In the second session, you will basically be doing 3 computerised tasks and we want to see if you become better with practice. Many things influence people's cognitive skills, including transient emotional states that the person may not even be particularly aware of. Because emotion changes frequently over periods of even an hour, emotion-relevant information will be collected periodically while you are working on the task. This information will permit our research team to control for the effects of these small shifts in emotion in analyzing the data. Furthermore, because the cardiovascular measurements are very sensitive, it is crucial that you refrain from talking, once you have started the computerized tasks until the end of the experiment. Diane will tell you exactly when this happens. ('No Talking' signs were also displayed around the computer cubicle to emphasize this point.)

Experimental manipulation: Guilt condition

After the research assistant attached all the sensors necessary for physiological measurements, she offered the participant more money than was originally agreed upon:

Research assistant: You know, these tasks actually take quite long - I'm going to give you R60 today. Some of the other participants haven't showed up, and I'm sure Haidee has loads of money anyway. I'm not really supposed to give you more than R30, but I'm sure it will be fine – please just don't mention this to anyone, OK?

The research assistant then briefly explained the computerized tasks to the participant, and reminded the participant that any form of verbal communication will be prohibited during the rest of the experiment to ensure signal integrity. She also instructed participants to sit quietly and relax after tasks were completed, and that she might go to the bathroom, but would be right back.

Participant performed RUN 1

Participant performed RUN 2 (The research assistant left halfway through RUN 2 to "go to the bathroom.")

The experimenter and supervisor entered unexpectedly as the participant finished RUN 2, with the purpose of checking up on something.

Supervisor (to participant): Sorry to interrupt… Melike, where is Diane?

Experimenter: I don't know, she probably just went to the bathroom.

Both the supervisor and experimenter then went behind the divider.

Supervisor: Melike, why don't you set up the next block of tasks for the participant in the meantime. (She finds the money box and discovers that money is missing.) No…this isn't right, something is wrong here…Melike can you come over here for a second?

Experimenter left the participant before the next block of tasks was properly set up, and went back behind the divider.

Supervisor: There is some money missing here - do you know anything about it?

Experimenter: No, I don't work with the money, Diane does.

Supervisor: How is this possible? I put enough money in here for at least 5 participants and now it is almost all gone! This isn't right, are you sure you don't know anything about it?

Experimenter: I'm sure. Diane woudn't just have just taken the money for herself, though. I think she probably pays participants more than she is supposed to.

Supervisor: Are you sure? Wait, lets find out…

The supervisor then went over to the participant and directly asked her whether the research assistant (Diane) had given her more than R30. (Participant nodded.)
Supervisor

(to experimenter): Well, this is just the last straw, and it has happened more than once now. Diane is just too nonchalant about this whole project… where is she anyway? I'm going to have to let her go. Please will you send her to me and start looking for a new assistant straight away. Actually…wait, you stay here and finish the participant and I'll go and find her.

Experimenter: Yes of course, I'm really sorry about it, I didn't know...

The experimenter then finished setting up the next block of tasks (RUN 3) for the participant, telling her not to worry and to keep concentrating on the tasks.

Experimental manipulation: Pride condition

A similar procedure was followed as for the Guilt condition, however, the research assistant did not offer participants any extra financial compensation.

Participant performed RUN 1

Participant performed RUN 2 (The research assistant left halfway through RUN 2 to "go to the bathroom.")

The experimenter and supervisor entered unexpectedly as the participant finished RUN 2, with the purpose of checking up on something.

Supervisor (to participant): Sorry to interrupt… Melike, where is Diane?

Experimenter: I don't know, she probably just went to the bathroom.

Both supervisor and experimenter then went behind the divider.

Supervisor: Melike, why don't you set up the next block of tasks for the participant in the meantime. (Supervisor is busy on the 'dummy computer' and struggles to find the right file.) Melike, can you come over here for a second?

Experimenter left the participant before the next block of tasks was properly set up, and went back behind the divider.

Supervisor: I can't find the test results for today – can you show me please?

Experimenter: Yes, sure – here they are. (She finds the right file.) So, for example, these are the previous participants' scores – you can see the data transfer is working quite nicely. Most of them do fine, but they seem to struggle with the first task. And these are 'Tess's" (current participant name) scores.
Supervisor: These scores are rather remarkable, aren't they?

Experimenter: Wow – yes I haven't seen them yet, but she really does seem to be outperforming almost everyone else. Nobody has done this well before.

Supervisor: And her reaction times are pretty fast as well. Look, she's just above the 95^{th} percentile! I'm really happy with these results and you guys are doing a great job. Wait, let me go tell her myself…

The supervisor then went over to the participant where she warmly congratulated her and smiled approvingly. (Participant smiled back in recognition.)

Supervisor (to experimenter): Do you mind finishing up here, I actually need Diane for something else.

Experimenter: Yes of course.

The experimenter then finished setting up the next block of tasks (RUN 3) for the participant, telling her that she was really impressed with her performance.

Experimental manipulation: Neutral condition

A similar procedure was followed as in the other two conditions, however, the research assistant did not offer participants any extra financial compensation, and no comments were made about the participant's performance.

The experimenter and supervisor again entered unexpectedly as the participant finished RUN 2, with the purpose of checking up on something on the computer.

Supervisor (to participant): Sorry to interrupt… Melike, where is Diane?

Experimenter: I don't know, she probably just went to the bathroom.

Both the supervisor and experimenter then went behind the divider.

The supervisor then started looking for a PowerPoint presentation on the 'dummy computer', which she ostensibly needed for a presentation the following day. The experimenter helped her to find this file and explained some concepts in the presentation. (The dialogue was timed to be of similar length as in the other conditions, and the content thereof was neutral.)

On their way out, the supervisor asked the participant if everything was going well. (Participant nodded.) She then asked the experimenter (Melike) to stay behind and finish the experiment because she needed the research assistant (Diane) for something else.

Appendix D: fMRI Pilot Study

In order to test the fMRI protocol, a behavioural pilot study was developed and conducted in the UCT Department of Psychology as an independent Honours project. The pilot study's main purpose was to confirm and optimize the validity of the IAT prejudice paradigm as a successful technique to elicit guilt and pride. It furthermore served to refine participant selection criteria for the larger study.

In short, a web-based screening survey was employed to recruit a large number of female participants ($N = 147$). From the survey, a subset of 19 participants who met all screening criteria participated in the study. Aside from certain demographic criteria, inclusion criteria involved selecting participants with a positive attitude toward all social groups included in the study (e.g., Black people, homosexual people, and Jewish people), as well as high social desirability concern.[10] During the testing session, participants also completed three other individual difference personality measures, namely the PANAS (Watson et al., 1988), the Affect Intensity Measure (AIM; Larsen & Diener, 1987), and the BIS/BAS scales (Carver & White, 1994). These were included to explore possible relations between the constructs assessed by these questionnaire measures and self-reported affect after the emotion manipulation.

Data from the pilot study confirmed that successful moral emotion elicitation is possible using preprogrammed IAT feedback. Self-report data as well as post-experimental interviews confirmed that target emotions of guilt and pride were differentially elicited by the IAT paradigm. The behavioural study furthermore served to identify types of individuals likely to experience intense target emotions. Of significance was the finding that BIS sensitivity was significantly correlated with self-reported pride ($r_s = .61$, $p < 0.01$) and satisfaction ($r_s = .54$, $p < .05$) in the Positive IAT (pride) condition. Pride and satisfaction were also highly correlated with each other in this condition ($r_s = .91$, $p < .001$). In the Negative IAT (guilt) condition, BIS sensitivity was also significantly correlated with self-reported guilt ($r_s = .53$, $p < .05$), as well as shame ($r = .50$, $p < .05$), but not embarrassment ($r_s = .27$, $p = .27$). In this condition, guilt was significantly correlated with shame ($r_s = .82$, $p < .001$), and to a lesser extent with embarrassment ($r_s = .53$, $p < .05$), while shame and embarrassment were highly correlated ($r_s = .83$, $p < .001$).

The above findings from correlational analyses were in line with data from Study 1, and suggested that participants with high BIS sensitivity are more likely to experience guilt and pride

[10] Specific details of the pilot selection procedures are not discussed here in full, because a comprehensive description of screening procedures is provided in Study 2 (the fMRI study).

during the Negative and Positive IAT conditions, respectively. None of the other personality measures were significantly related to self-reported affect, except the Positive Affect (PA) scale, which showed a significant negative correlation with guilt ($p < .01$), as well as shame ($p < .05$).

Some participants in the pilot study did not report experiencing target emotions, however, especially during the Negative IAT condition. Post-experimental interviews suggested that these individuals were often personally involved in cross-racial or gay relationships, or had close relatives or friends in these categories. They therefore did not believe Negative IAT feedback suggesting that they were prejudiced against Black or gay individuals. In order to avoid similar confounding effects in future, it was thus considered important to include indications of homosexual affiliations (e.g., “Do you have close gay friends?”), as well as cross-racial contact (e.g., “Do you have many Black friends/family?”) in future screening measures.

Overall, the pilot study suggested that the elicitation of specific moral emotions using an IAT prejudice paradigm with preprogrammed feedback was reliable, yet highly dependent on participant selection criteria. High BIS-sensitive individuals furthermore appeared to be more likely to experience both guilt and pride using this paradigm.

Appendix E: Web-based Survey (Demographic Section)

UCT Social Preference Survey

We are conducting research on social 'prejudice' or 'preference' tendencies among students at UCT. Please complete the questionnaires below, which shouldn't take you more than roughly 30 minutes. In return for your participation (and completion of *all* forms) you will receive 2 units toward your SRPP credits. Your responses will be kept completely confidential.

There are no right or wrong answers, so please respond as honestly as possible.

You are first required to fill out some demographic details on the next page. Please note that we need your correct student number in order to award you course credit. All personal information will be kept separate from your survey responses, so that those survey responses remain completely confidential.

Student number: ___________________

Gender: __________________________

Home language: ___________________

Handedness: ______________________

Race:
Please select one of the following: Asian/Indian, Black, Coloured, or White

Age: (fill in)

Religious Denomination: (fill in)

Attitude toward religion in general:
Please select one of the following: Positive, Negative, or Neutral

Sexual Orientation:
Please select one of the following: Homosexual, Heterosexual, or Bisexual

Please give an indication of the following in the space provided. Please answer all questions:

1) Have you ever dated someone of the opposite race?

2) Do you have close gay or lesbian friends?

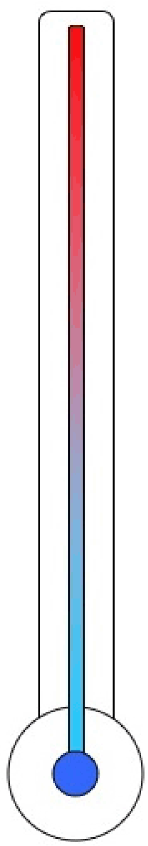
3) Do you have close friends or family of a different race?

Appendix F: Rating Thermometer

We are interested in people's personal attitudes toward a variety of stereotyped social groups.

Below you will see something that looks like a thermometer. You will be using this to indicate your overall evaluation of these social groups.

Here's how it works: If you have a favourable attitude toward members of a group, you would give the group a score somewhere between 50° and 100°, depending on how favorable your evaluation is. On the other hand, if you have an unfavorable attitude toward members of this group, you would give the group a score somewhere between 0° and 50°, depending on how unfavorable your evaluation is. The degree labels will help you to place your evaluation. However, you are not restricted to the numbers indicated, feel free to use any number between 0 and 100.



100° Extremely favourable

90° Very favourable

80° Quite favourable

70° Fairly favourable

60° Slightly favourable

50° Neither favourable nor unfavourable

40° Slightly unfavourable

30° Fairly unfavourable

20° Quite unfavourable

10° Very unfavourable

Please give your thermometer rating for Black people in the space provided: ________________

Appendix G: MRI Screening Form

Cape Universities Brain Imaging Centre (CUBIC)

MRI Volunteer Screening Form

Volunteer Information:

Name	Contact number
Date of Birth	Project name
Weight	Principle investigator

The following information is very important to ensure your safety and to prevent any interference during the MRI procedure.

Please answer the following questions (mark with an X):	**Yes**	**No**	**Don't know**
Pacemaker			
Aneurism clips			
Artificial heart valve			
Vena cava filter			
Prosthesis (e.g. eye, breast etc)			
Shrapnel in eye or body			
Neurostimulator			
Cochlear implant (ear) or hearing aid			
? Diabetic			
? Renal impairment			
? Asthma			
? Allergies			
? Any other implants (e.g. screws, plates, joint replacements)			
? Pregnant			
? Previous MRI investigation with intravenous contrast			
Is there any other device implanted or are there any other ailments that you think that we should be aware of?			

I hereby acknowledge that the potential risks of the examination have been explained to me and that during the course of the investigation it may be necessary for the intravenous injection of a contrast agent.

Attention: It is the policy of this institution not to discuss results of the MR investigation with the patients for ethical reasons. All enquiries in this regard should be directed to the referring physician.

Signature: **Date:**

Cape Universities Brain Imaging Centre
Fisan Building, Faculty of Health Sciences, University of Stellenbosch, Tygerberg, 7505
Tel: 27-21-938-9646 Fax: 27-21-938-9728 www.sun.ac.za/cubic

Appendix H: fMRI Consent Form

UNIVERSITY OF CAPE TOWN
DEPARTMENT OF PSYCHOLOGY

fMRI Participant Information Leaflet
and Consent Form

Title of Research Project: Functional brain imaging of the Implicit Association Test (IAT) in healthy low-prejudice individuals.

Reference Number: 247/2009

Principal Investigator: Melike Fourie

Contact Details:
Department of Psychology
University of Cape Town
Office: 021 – 650-3415
Mobile: 083 781 6021

Supervisor: Dr Kevin G.F. Thomas

Contact Number: 021-650-4608

Dear Volunteer

You are being invited to take part in a research project. Please take some time to read the information presented here, which will explain the details of this project. The Principal Investigator (the person in charge of this research) will describe this study to you and answer all of your questions. It is very important that you are fully satisfied that you clearly understand what this research entails and how you could be involved. Also, your participation is entirely voluntary and you are free to decline to participate. If you say no, this will not affect you negatively in any way whatsoever. You are also free to withdraw from the study at any point, even if you do agree to take part.

This study has been approved by the UCT Psychology Ethics Committee, as well as the Research Ethics Committee of UCT Faculty of Health Sciences and will be conducted according to the ethical guidelines and principles of the international Declaration of Helsinki, South African Guidelines for Good Clinical Practice and the Medical Research Council (MRC) Ethical Guidelines for Research.

Research Objectives

The main purpose of this research essentially involves identifying, through the technique of fMRI, which areas of the brain are active or take part in performing the Implicit Association Test (IAT). The IAT is a well-documented reaction-time test that gives a reliable indication of your unconscious bias or prejudice on things like race, age and gender. Specifically, we are interested in identifying brain areas associated with non-prejudiced performance.

Study Procedures

This study essentially involves two sessions, both of which will be conducted by the principal investigator listed above. During the first session you will be asked to complete some questionnaires assessing general mood as well as medical history (e.g., previous neurological or psychiatric illness), as well as an MRI suitability questionnaire which will mainly involve screening for pregnancy and any metal objects/implants in your body. If you meet the inclusion criteria and agree to participate, you will be contacted and scheduled to have a type of brain scan, called an fMRI (functional magnetic resonance imaging) scan. This session shouldn't take more than 30mins.

For this second session you will be required to go to Tygerberg Hospital, where the fMRI scanner is located. Two qualified radiographers will take care of all the scanning and MRI safety procedures, such as making sure that the scanning room stays free of magnetic material. They will also perform a formal screen to make sure that it is completely safe for you to have a scan. If transport is a problem, a lift will be arranged for you from UCT to Tygerberg Hospital. During the scan, you will be asked to perform various versions of the IAT, which will enable the investigators to determine your brain function. Performing the IAT is relatively simple and basically involves sorting stimuli through button presses with either your right or left index fingers. You will be allowed to practice the task first outside of the scanner to become familiar with the instructions. In addition, during the scan some physiological sensors will be attached to your body to measure your heart rate during the tasks. It is estimated that the entire procedure, from once you arrive at the imaging centre (CUBIC) until when you leave, will not exceed roughly 2hours.

As the scan is done in a relatively confined space, occasionally people become anxious. This does not happen often, and if you feel anxious, we will spend time allowing you to get used to the surroundings. The scan will require you to lie on your back on a table that will move into the scanning machine for the 40 minutes it will take for the scan. During this time you will be able to rest between different IATs. You will also be able to talk to the study doctor/assistant at all times during the scan. The scan is a safe procedure if you have been screened correctly for the presence of any magnetic material on or inside you such as pace-makers, surgical clips and metal objects in the eyes. When the machine performs scans, you will hear loud banging noises, but you will be clearly warned when this will take place. At this time you will feel nothing and the noise is not harmful to you in any way. To minimise the possible discomfort associated with this, we will give you some soft earplugs to put in and we will also give you earphones to wear during the scan.

After the experimental session is over, you will be informed in detail about the design of the study and the research questions we hope to answer. You will also have the opportunity to ask questions and thus learn more about psychological research.

MRI Information

Safety and Metal Objects: The MRI scanner is a powerful magnet. Because metal objects are strongly attracted to the magnet, any metal objects you are carrying or wearing must be removed prior to the MRI scan to avoid potentially severe injury

Unexpected Findings: In rare cases, researchers discover unexpected findings related to a participant's MRI scan in which case the scan is referred to a radiologist for further analysis. Further tests may be recommended in order to determine the nature and significance of the unexpected finding, in which case that participant will be referred to a General Practitioner of their choosing.

Ongoing disclosure of potential harms: If new findings about the potential harms of the MRI technique become available during the time of the study, the researcher will inform you.

Discomfort Associated with the Study

There are only low or minimal risks associated with your participation in this study. If you feel tired at any point in either of the visits, you should please ask the principal investigator for a rest. If for some reason you are unable to complete a visit on a particular day we may reschedule to complete the assessments at another time. If you wish to discuss the information above or any discomforts you may experience, you may ask questions now or call the Principal Investigator listed on the front page of this form.

Potential Benefits

There may be no direct benefits to you for participating in this study. However, you will be making an important contribution to research that may benefit others in the future. This research will allow us to gather information about the performance of healthy adults on the administered tests.

Compensation for Study Participation

You will receive financial compensation of R150 for any cost incurred in attending the prescribed study visit. In addition, you will receive a compact disc with a 3D scan of your own brain.

Confidentiality

Your participation is regarded as strictly confidential. The results of the study will be published in the professional literature and form part of a larger PhD thesis, but your identity will not be revealed at any time to people outside of the study team.

The right to ask questions and withdraw from the study

You have the right to ask questions at any time about any aspect of the study. If you have any queries, you can contact:

Melike Fourie: Tel (24hr contact number): 083-781 6021
Dr Kevin Thomas: Tel (office hours): 021-650-4608
Prof Ernesta Meintjes Tel (office hours): 021-406 6547

If you have any questions about your rights as a research participant, you may contact the Chairperson of the Research Ethics Committee, Prof Marc Blockman (Tel: 021-406-6496).

Your participation in the study is entirely voluntary. You have the right to withdraw at any time. If you decide to withdraw from the study, it will not jeopardize you or any future treatment you may require in any way.

You are entitled to a signed copy of this document.

If you agree to take part, please complete the following section.

I (name)………………………………………………. have been invited to take part in the above research project entitled *Functional brain imaging of the Implicit Association Test (IAT) in healthy low-prejudice individuals.*
The principal investigator/study doctor has explained the details of the study to me and I understand what they have said to me.
I also know that I am free to withdraw from the study at any time if I am unhappy.
By writing my name below, I voluntary agree to take part in this research project. I confirm that I have not been forced in any way or by anyone to take part.

____
Name of Participant (printed)

__________________________ ________________
Signature of Participant Dated

Declaration by investigator:

I (name) ……………………………………………..……… declare that:

I explained the information in this document to ………………………………………
I encouraged him/her to ask questions and took adequate time to answer them.
I am satisfied that he/she adequately understand all aspects of the research, as discussed above.

Signed at (place) ....…........…………….. on (date) …………...………...

____
Signature of investigator

Appendix I: Affective Ratings of fMRI IAT Stimuli

The primary aim of this study was to obtain affective ratings for all stimuli that were selected to be programmed as part of the fMRI protocol. For the fMRI study, six novel IATs were necessary, each consisting of various images and attributes as well as pre-programmed response feedback. Although emotion elicitation was entirely dependent on the response feedback, rather than the actual IAT stimuli, it was nonetheless important that IAT stimuli not elicit any strong emotions (either positive or negative) with the potential to interfere with desired emotional responses. By obtaining psychometric properties of all picture stimuli used, I aimed to validate their use in the fMRI study.

To ensure that judgments for my target pictures were sufficiently neutral, they were compared to both positive and negative emotional pictures from the International Affective Picture System (IAPS; Lang, Bradley, & Cuthbert, 1999). The IAPS is a normative set of colour images specifically developed for emotion and attention research; the pictures in the set cover a wide variety of semantic categories. Each picture in this set had been standardized through ratings obtained from hundreds of participants using the Self-Assessment Manikin (SAM; Lang, 1980). This affective rating scale records emotional assessments of stimuli in three dimensions, based on Osgood's factor analysis of the semantic differential scales (Osgood, Suci, & Tannenbaum, 1957). The two principle dimensions are affective valence and arousal. A third dimension, dominance or control, typically explains a smaller amount of the variance in affective ratings. By using the same normative rating procedure that was used for the IAPS, I was able to compare ratings obtained for my picture set to that of the established normative picture set. In addition, this strategy allowed me to determine whether participants were performing the rating task appropriately.

Method

Participants

Forty-three female participants between the ages of 18 and 25 years (*M* = 19.87, *SD* = 2.38) were recruited from the UCT undergraduate population through the Department of Psychology's Student Research Participation Program. All received course credit in exchange for participation in the 30 min study. Two participants' data were discarded because they did not follow instructions; the final sample thus consisted of 41 participants.

Stimuli

Six IATs were developed for the three fMRI stimulus conditions: Neutral, Positive, and Negative. Each of these IATs involved two contrasted concepts (e.g., Black-White, or homosexual-heterosexual). Based on Nosek and colleagues' (2005) findings and recommendations in constructing IATs, I decided to use four stimulus items per concept (i.e., 8 images for each IAT). A total of 48 pictures were therefore sourced from the Internet, as well as from the NimStim set of facial expressions (Tottenham et al., 2009). The picture set ranged across topics for the 6 different IATs, namely facial hair, glasses, race, disability, sexuality and religion.

Human faces were employed as a common denominator in all IATs to avoid fMRI artifact due to facial recognition in some, but not in other, conditions (Kanwisher, McDermott, & Chun, 1997; Rotshtein, Henson, Treves, Driver, & Dolan, 2005). Faces were specifically selected to appear as close to neutral in expression as possible, and were matched across IAT categories for attractiveness and age. To promote uniformity among the different IAT conditions, and to prevent gender biases from interfering with participant responses, all selected stimuli were male where possible (only two images included females, both as part of a heterosexual couple in the sexuality IAT). Therefore, the facial hair IAT featured male faces with or without beards or mustaches; the glasses IAT featured male faces with or without spectacles; the race IAT featured Black and White male faces; the disability IAT featured images and symbols of disabled and healthy/normal males; the sexuality IAT featured images and symbols of heterosexual couples and homosexual male couples; and the religion IAT featured images and symbols of Jewish males, and males practicing other religions including Christianity, Islam, Hinduism, and Buddhism.

For comparison purposes, 10 positively- and 10 negatively-valenced pictures were selected from the IAPS. The average valence, arousal, and dominance values for these stimuli, based on previous SAM ratings by female participants (Lang, Bradley, & Cuthbert, 2005), are presented in Table I1.

Table I1

IAPS Stimuli Normative Ratings of Valence, Arousal, and Dominance: Female Subjects

Description	Picture No.	Valence Mean (*SD*)	Arousal Mean (*SD*)	Dominance Mean (*SD*)
Practice Items				
1 Cow	1670	5.88 (1.84)	3.52 (2.05)	5.40 (1.74)
2 Fire	9921	1.58 (1.06)	6.87 (1.94)	3.56 (1.60)
3 Woman	2030	6.02 (1.48)	3.08 (1.72)	5.31 (1.59)
Positive Items				
1. Women	1340	7.63 (1.52)	5.25 (2.24)	5.85 (1.75)
2. Puppies	1710	8.59 (0.99)	5.31 (2.54)	6.50 (2.09)
3. Bunnies	1750	8.59 (0.75)	4.02 (2.40)	6.28 (2.01)
4. Porpoise	1920	7.94 (1.61)	4.31 (2.57)	6.56 (2.05)
5. Mickey	1999	7.68 (1.52)	5.02 (2.48)	6.85 (2.00)
6. Baby	2040	8.74 (0.64)	4.97 (2.85)	7.32 (1.91)
7. Baby	2070	8.50 (1.28)	4.84 (2.97)	7.44 (1.96)
8. Giraffes	1601	7.17 (1.50)	4.02 (1.97)	6.51 (1.89)
9. Adult	2010	6.77 (1.86)	3.74 (2.14)	5.85 (1.77)
10. Adult	2000	7.10 (1.62)	3.72 (2.31)	6.10 (1.57)
Average		**7.87 (0.71)**	**4.52 (0.62)**	**6.53 (0.55)**
Negative Items				
1. Snake	1019	3.46 (2.03)	6.06 (1.83)	4.03 (2.10)
2. Spider	1201	2.93 (1.81)	6.87 (2.09)	3.82 (2.26)
3. PitBull	1300	3.41 (1.63)	6.70 (2.04)	3.34 (2.27)
4. Shark	1930	3.56 (1.90)	6.71 (1.91)	2.93 (1.91)
5. Baby	2053	2.17 (1.90)	5.83 (2.38)	3.40 (2.25)
6. Heroin	9102	3.03 (1.75)	5.13 (2.53)	4.35 (2.42)
7. Cemetery	9220	1.86 (1.46)	4.16 (1.84)	3.00 (1.72)
8. Garbage	9290	2.76 (1.44)	4.44 (2.01)	5.11 (2.08)
9. Dental Exam	9582	4.24 (2.64)	5.38 (2.36)	4.44 (2.32)
10. Mutilation	9253	1.60 (0.99)	5.65 (2.58)	3.48 (2.10)
Average		**2.90 (0.83)**	**5.69 (0.94)**	**3.79 (0.70)**

Note. These data were taken from the Instruction Manual for the IAPS (Lang et al., 2005). Ratings range from 1 to 9, with 9 signifying a high rating on each dimension.

Affective Rating Scale

The Self-Assessment Manikin (SAM; Lang, 1980) is a nonverbal self-report scale that provides a relatively quick and easy method for obtaining affective assessments of picture stimuli. In my study, a paper-and-pen version of the SAM was employed, which makes use of graphic figures to depict the three main affective dimensions of pleasure (or valence), arousal and dominance (see Figure I1).

The pleasure dimension of SAM ranges from a happy cheerful manikin to a sad and unhappy one, while the arousal dimension ranges from an excited wide-eyed manikin to one that appears sleepy and with closed eyes. The dominance scale represents changes in control, and ranges from a very small manikin (representing a feeling of being controlled or submissive), to a very large manikin (representing a feeling of being in control or powerful). For each of these dimensions, respondents are required to mark their assessment of a stimulus by placing an 'X' over any of the five manikins of a dimension, or between any two to make a more finely graded rating. The SAM therefore comprises a 9-point rating scale, such that 9 signifies a high rating on each dimension (i.e., high pleasure, high arousal, and high dominance), and 1 signifies a low rating on each dimension (i.e., low pleasure, low arousal, and low dominance). (In order to obtain these ratings, the pleasure and arousal scales have to be reversed-scored.)

Design and Analysis

Participants were run in groups of two in two separate venues. The 68 pictures that were rated were arranged in a semi-random sequence in Microsoft PowerPoint, with the constraint that no more than 2 pictures from the same stimulus condition were presented consecutively. Pictures were displayed for 5 s on a 13 inch computer monitor, situated approximately 0.5 m from the participant. Four different picture orders were created to ensure that the position of a particular image was counterbalanced within the series of pictures. Every 5th participant therefore viewed the same sequence of images. The size of each image on screen was roughly 10 cm x 12 cm, which was similar to the final dimensions used for stimuli in the fMRI protocol.

The three dimensions of SAM (i.e., pleasure, arousal, and dominance) served as the dependent measures in data analysis. For all repeated-measures analyses, the multivariate test statistic (Pillai's trace) is reported to avoid potential sphericity issues (O'Brien & Kaiser, 1985).

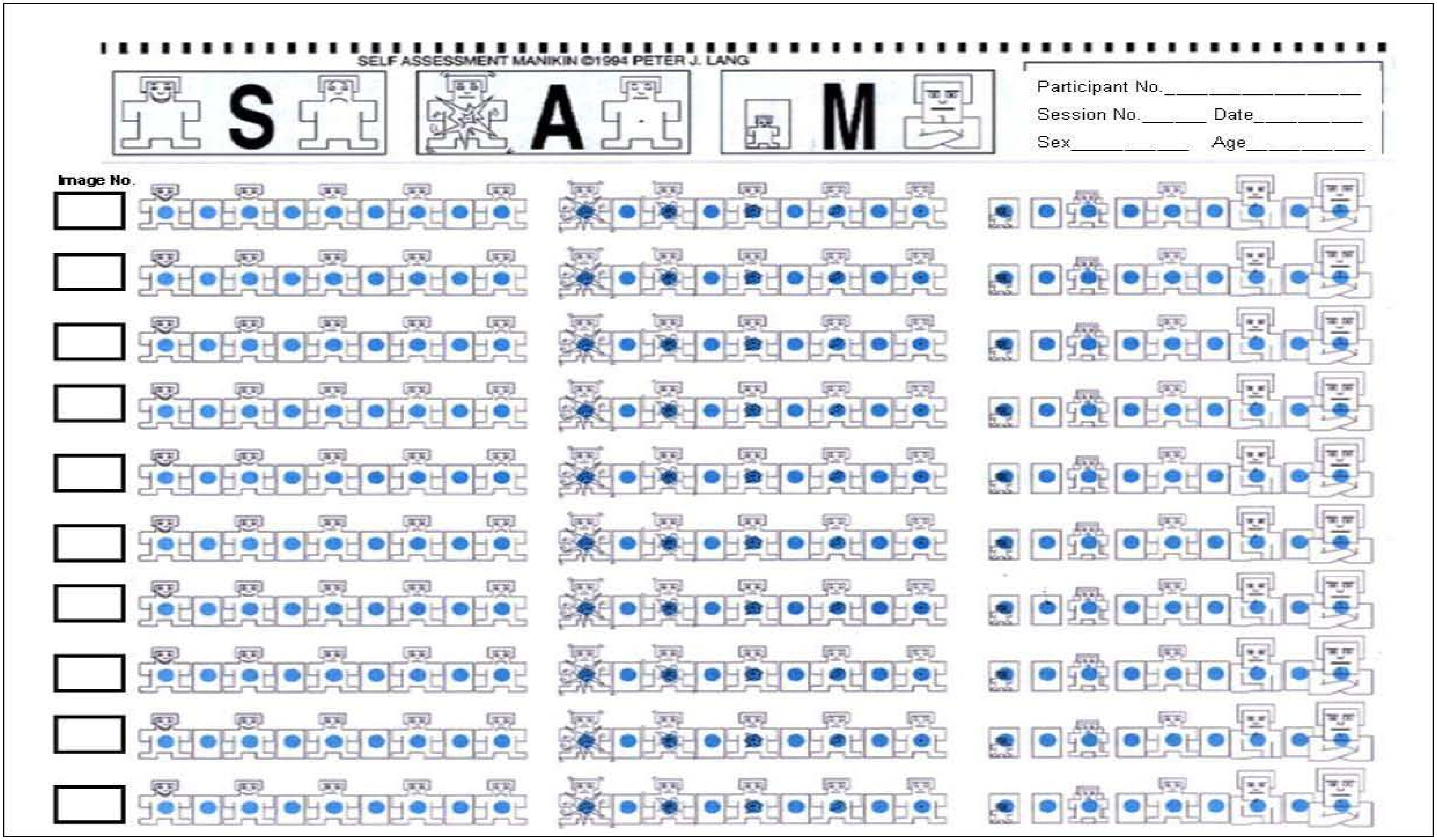


Figure I1. The Self-Assessment Manikin (SAM) answer sheet.

Procedure

On arrival, the purpose of the study was explained, and informed consent and biographical details were obtained from each participant. The SAM rating procedure, which was similar to the normative rating procedure of the IAPS (Lang et al., 2005), was then thoroughly explained to each participant by demonstrating ratings in an answer booklet.

In short, the pleasure dimension was explained to participants as ranging from happy (smiling manikin) to unhappy (frowning manikin). They were told that the one end of the scale should be selected (by placing an “X” over the figure) when one feels happy, pleased or satisfied, while the other end of the scale should be selected when one feels unhappy, annoyed or unsatisfied. They were also told that intermediate feelings of pleasure should be indicated by selecting manikins between the two ends, while the manikin in the middle of the scale should be selected if one feels completely neutral, i.e., neither happy nor sad. The arousal dimension was explained as ranging from feeling excited (manikin with eyes wide open) to calm (manikin with eyes closed). At one end of this scale one feels aroused, stimulated or jittery, while at the other end one feels completely relaxed, calm or dull. Participants were again encouraged to select figures in-between for more finely graded ratings that described their feelings more accurately. Finally, the dominance scale was explained as ranging from controlled (small manikin) to in-control (large manikin). Participants were told that at one end of this scale one feels submissive, influenced or controlled, while at the other end of the scale one feels dominant, autonomous or in-control. Participants were instructed to rate pictures as honestly as possible and that the goal of the study was to determine how each picture made them feel while viewing it.

Following these instructions, participants viewed and rated three practice pictures, as in the IAPS normative procedure (see Table I1). The purpose of the practice pictures were to ensure that all instructions were understood correctly and to help anchor the emotional rating scales. A positive, negative, and neutral image that was not used in the study, were thus presented. At this stage participants were allowed to ask questions of the experimenter (either myself or a research assistant). Once the experimenter was satisfied that all rating instructions were understood, the experimental session commenced.

Each participant was seated in front of a computer monitor and instructed to view each picture for the entire time it was screened (5 s), and to then make their ratings of all three affective dimensions in their own time. When finished, they could move on to the next picture by pushing the ‘down’ arrow on the keyboard. Each picture was numbered and corresponded with a particular number in the answer booklet to avoid confusion. Following the presentation of the last picture, answer booklets were collected and participants were debriefed.

Results

Participant Demographics

The sample of participants ($N = 41$) who successfully completed all rating procedures could be divided into the following racial demographic groups: White (61%); Coloured (22%); Black (9.7%); and Asian/Indian (7.3%). Because I only intended to scan White participants in the fMRI study, it was important to check for demographic group differences in affective ratings that may cause the data to be unrepresentative of my target sample. Toward this purpose, the group was broadly divided into White ($n = 25$) and Other ($n = 16$) participants. I tested for differences in SAM ratings between these groups by performing independent-samples *t*-tests on the pleasure, arousal, and dominance dimensions of the positive and negative IAPS picture stimuli (Table I2). None of these analyses reached significance (*p*s > .2), although there was a slight tendency for White participants to experience higher arousal than Other participants for the negative IAPS stimuli ($p = .09$).

Table I2

Pleasure, Arousal, and Dominance Values for IAPS Stimuli, as Rated by Other (n = 16) and White (n = 25) Participant Groups

		Pleasure/Valence Mean (*SD*)	Arousal Mean (*SD*)	Dominance Mean (*SD*)
IAPS Positive:	Other	7.49 (0.75)	6.53 (1.14)	6.74 (1.14)
	White	7.43 (0.52)	6.19 (1.00)	6.35 (0.92)
IAPS Negative:	Other	2.37 (0.88)	5.69 (1.40)	2.82 (1.22)
	White	2.19 (0.60)	6.38 (1.10)	2.37 (1.17)

Because no significant differences existed between White and Other participant groups on measures of pleasure, arousal, or dominance, this demographic grouping was not employed in subsequent analyses. The results reported below therefore represent data from all racial groups.

Ratings of Valence, Arousal, and Dominance

The means and standard deviations of affective ratings for all picture stimuli can be seen in Table I3. The results are presented according to picture group classification: IAPS Positive, IAPS Negative, IAT Positive, IAT Negative, and IAT Neutral. Because SAM records affect on a 9-point rating scale for each dimension, a rating of 5 is considered neutral.

From the data it is evident that participants rated the positive IAPS pictures as fairly pleasant, while the negative IAPS pictures were consistently rated as more negative or unpleasant. The IAPS values obtained in this study were furthermore very similar in valence to the normative averages for these pictures, while all ratings were slightly higher in terms of arousal values (see Table I3). By comparison, the ratings for stimuli from the different IAT stimulus conditions (i.e., Positive, Negative, and Neutral) were clustered around 5 for all three affective dimensions, and could therefore be regarded as fairly neutral.

Table I3

Affective Ratings for the Different Stimulus Groups, Assessed With the Self-Assessment Manikin

Stimulus group	Pleasure/Valence Mean (SD)	Arousal Mean (SD)	Dominance Mean (SD)
IAPS Positive (*n* = 10)	7.45 (0.61)	6.32 (1.06)	6.50 (1.02)
Normative values	***7.87 (0.71)***	***4.52 (0.62)***	***6.53 (0.55)***
IAPS Negative (*n* = 10)	2.26 (0.72)	6.11 (1.26)	2.55 (1.20)
Normative values	***2.90 (0.83)***	***5.69 (0.94)***	***3.79 (0.70)***
IAT Positive (*n* = 16)	5.79 (0.74)	5.38 (0.86)	5.53 (1.06)
IAT Negative (*n* = 16)	4.88 (0.44)	5.02 (0.77)	5.27 (1.02)
IAT Neutral (*n* = 16)	5.15 (0.50)	5.09 (0.82)	5.42 (0.96)

Note. Normative values reflect ratings for the selected IAPS stimuli, as taken from the IAPS Instruction Manual (Lang et al., 2005).

Comparison of Positive IAPS, Negative IAPS, and IAT Stimuli

To determine if there was an overall difference in affective ratings between the IAPS pictures and my IAT pictures, the three IAT stimulus conditions were clumped together (i.e., Positive, Negative, and Neutral IAT stimuli). The resulting three stimulus groups, namely IAPS

Positive, IAPS Negative, and IAT Together, were compared through one-way repeated measures analysis for valence, arousal, and dominance respectively (see Figure I2).

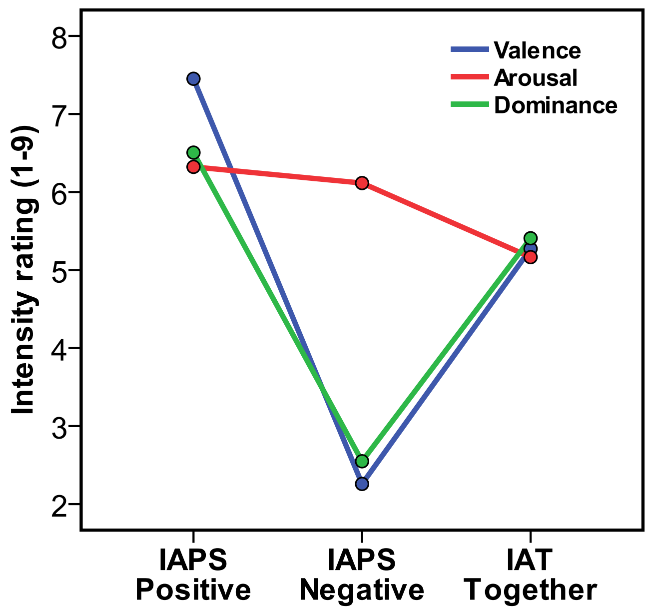


Figure I2. Intensity ratings across the three affect dimensions for the IAPS Positive, IAPS Negative, and IAT Together stimuli.

For valence ratings, the repeated-measures ANOVA was significant, $F(2,39) = 448.55$, $p < .001$, $\eta^2 = .96$ (Pillay's trace).[11] Post-hoc Bonferroni comparisons detected statistically significant differences between IAPS and IAT stimuli: While IAPS Positive stimuli were significantly more pleasant than IAT Together stimuli ($p < .001$, $r = .96$), IAPS Negative stimuli were significantly more unpleasant than IAT Together stimuli ($p < .001$, $r = .97$). For arousal ratings, the repeated- measures ANOVA also detected statistically significant differences between groups, $F(2,39) = 53.71$, $p < .001$, $\eta^2 = .73$ (Pillay's trace). Post-hoc Bonferroni comparisons indicated that IAT Together stimuli were significantly lower in arousal than IAPS Positive, as well as IAPS Negative stimuli (*ps* < .001, *rs* > .65), while the IAPS stimulus groups were of similar arousal values ($p = 1.00$). Lastly, the repeated-measures ANOVA for dominance

[11]MANOVA test statistics are reported for each analysis, because Mauchly's test indicated that the assumption of sphericity had been violated at the $p < .001$ level.

ratings was also statistically significant $F(2,39) = 158.22$, $p < .001$, $\eta^2 = .89$ (Pillay's trace); with post-hoc Bonferroni comparisons indicating significant differences between all stimulus groups ($p < .001$ for all three comparisons).

When the means of the valence and arousal ratings for all 68 pictures were plotted in two-dimensional affective space, a characteristic boomerang-shaped distribution for the stimuli could be detected (e.g., Bradley, Codispoti, Cuthbert et al., 2001; Lang et al., 1999, 2005)[12]. This shape can be attributed to the fact that stimuli judged as either very pleasant or very unpleasant are also usually more arousing than stimuli judged as more neutral. A graphic display of these data can be seen in Figure I3. The distribution clearly indicates that most IAT picture stimuli (with some exceptions) are centred at values of 5 for both valence and arousal dimensions.

From these results it could be concluded that my IAT stimuli were significantly different from the IAPS stimuli, thus neither positive nor negative, and more neutral in terms of valence, arousal, and dominance.

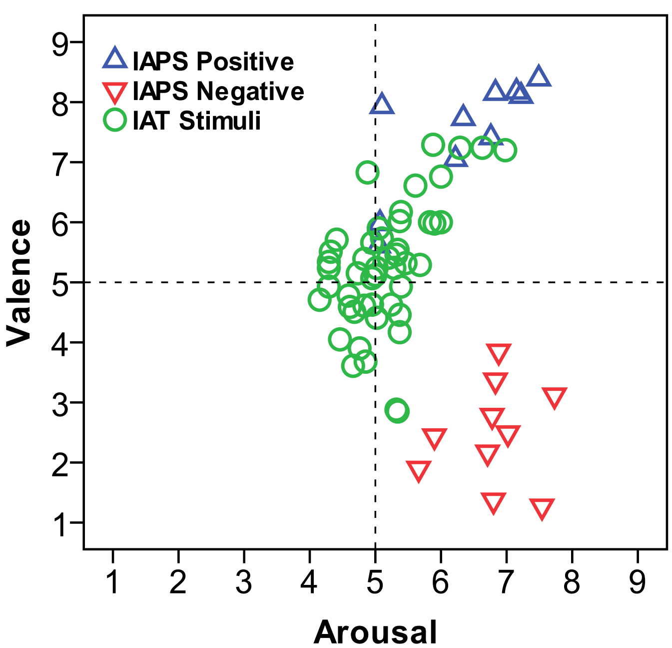


Figure I3. A two-dimensional plot of affective space, employing mean valence and arousal ratings of all 68 picture stimuli.

---

[12]Because the dominance rating typically explains a small amount of variance, this dimension was not plotted.

Comparison of IAT Stimulus Conditions

To determine more specifically whether differences existed between the different IAT stimulus conditions, IAT stimuli were again divided into their respective stimulus conditions, i.e., Neutral, Positive, and Negative. These groups were then compared through repeated-measures analysis for valence, arousal, and dominance respectively (See Figure I4)[13].

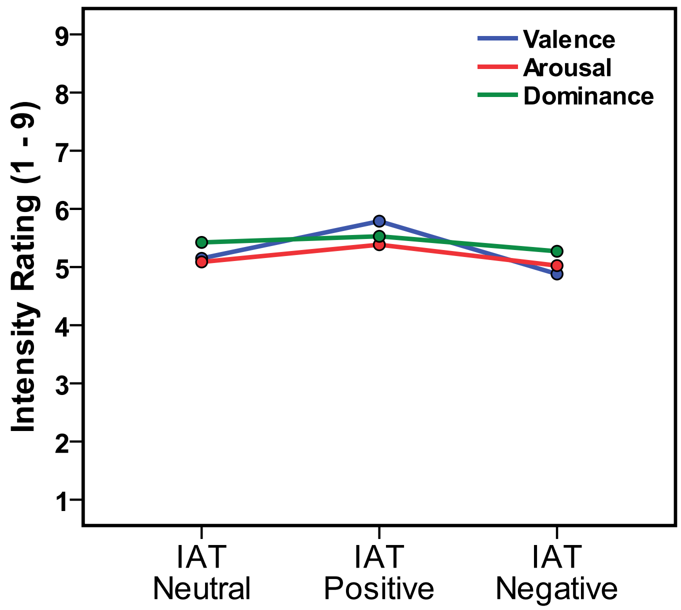


Figure I4. Affective intensity ratings for stimuli from the thee IAT stimulus conditions, i.e., Neutral, Positive, and Negative.

The one-way repeated-measures ANOVA for valence was significant, $F(2,39) = 28.73$, $p < .001$, $\eta^2 = .60$ (Pillai's trace), as was a similar analysis for arousal $F(2,39) = 4.20$, $p < .05$, $\eta^2 = .18$ (Pillai's trace). Bonferroni post-hoc comparisons indicated significant differences between IAT conditions for valence: IAT Positive stimuli were rated as significantly more pleasant than IAT Neutral ($p < .001$, $r = .65$), as well as IAT Negative stimuli ($p < .001$, $r = .76$), while IAT Neutral stimuli were also significantly more pleasant than IAT Negative stimuli ($p < .01$, $r = .53$). In terms of arousal, Bonferroni post-hoc comparisons indicated that the only significant difference existed between IAT Positive and IAT Negative stimuli, such that IAT Positive

[13]MANOVA test statistics are again reported for each analysis.

stimuli were rated as more arousing than IAT Negative stimuli ($p < .05$, $r = .42$). The one-way repeated-measures ANOVA for dominance between IAT stimulus conditions was not statistically significant ($p = .07$, $\eta^2 = .07$).

These results indicate that significant differences existed in terms of valence, and to a lesser extent arousal, between different IAT stimulus conditions. It should be noted, however, that IAT Positive stimuli were still significantly less pleasant than IAPS Positive stimuli, $t(40) = 11.39$, $p < .001$, $r = .87$, while IAT Negative stimuli were still significantly less unpleasant than IAPS Negative stimuli, $t(40) = -21.32$, $p < .001$, $r = .96$.

Discussion

The rationale for the current study was to determine whether stimuli selected for the various IAT stimulus conditions were suitable for use in the fMRI study. To this end, IAT stimuli had to be sufficiently neutral in terms of valence and arousal values so as to not elicit any unwanted emotional reactions during the imaging procedure. IAT stimuli, as well as positive and negative images from the IAPS, were consequently rated on measures of pleasure, arousal, and dominance using a tool commonly used in emotional assessment studies, namely the Self-Assessment Manikin (SAM). At a glance, the SAM ratings for my IAT stimuli were fairly neutral across all affective dimensions; however, a more detailed discussion of results is presented below.

Because no significant differences existed between participants' emotion ratings across different racial demographic groupings, it was considered acceptable to analyze data of all participants together. IAT stimuli were then compared as a group (i.e., Positive, Negative, and Neutral stimulus conditions clumped together) to both positive and negative IAPS stimuli. Results indicated that positive IAPS stimuli were rated as significantly more pleasant and arousing than IAT stimuli, whereas negative IAPS stimuli were rated as significantly more unpleasant and arousing than IAT stimuli. IAT stimuli were also rated as significantly more neutral than IAPS stimuli in terms of dominance. These results are consistent with a plot of all stimuli in two-dimensional affective space based on mean valence and arousal ratings: Whereas IAPS pictures tended to cluster in the quadrants of emotional space depicting high arousal and high positive or negative valence (i.e., pleasant and unpleasant material), IAT stimuli were clustered predominantly around neutral values for valence and arousal.

Comparison of IAT stimulus conditions with each other, however, resulted in significant differences. The dimension that distinguished IAT conditions most significantly was valence: IAT Positive stimuli were rated as more pleasant than IAT Neutral and IAT Negative stimuli, while IAT Negative stimuli were rated as more unpleasant than IAT Neutral stimuli. Although these differences were *statistically* significant, it is evident when one inspects Figure I3 that there is little *practical* significance in these findings – stimuli from all IAT conditions were clustered around relatively neutral valence values. Moreover, there were significant differences between the IAT Positive stimuli and the IAPS Positive stimuli in terms of valence, and similarly, between the IAT Negative stimuli and the IAPS Negative stimuli. Taken together, the IAT stimuli cannot be regarded as positive or negative *per se*, but only as positive or negative in relation to each other.

One may ask, however, how these differences in valence between different IAT stimulus conditions were brought about; why did IAT Positive pictures appear more pleasant and IAT Negative pictures appear more unpleasant? A closer inspection of specific valence ratings for each image provided some answers. In both the IAT Positive and Negative conditions, it appeared that certain images (about 4 per condition) were consistently rated as more pleasant or unpleasant, respectively. These images depicted people with disability and some Black individuals in the IAT Negative condition, and heterosexual couples and some religious symbols in the IAT Positive condition (see Figure I5). Participants therefore experienced any images or symbols illustrating disability as unpleasant, and there was also a tendency to rate Black faces as more negative than White faces, probably because the majority of participants were White themselves. Participants also rated any image illustrating heterosexual couples as high in pleasure, while the cross (the symbol for Christianity) was also consistently rated above average in terms of pleasure. Because of the nature of the IAT topics that were selected for the IAT Negative and Positive conditions (i.e., disability, race, sexuality, and religion), it was not feasible to find more suitable or neutral stimuli depicting the same subject matter than the images already employed.

The images in the IAT Neutral condition that were rated as either too pleasant (n = 2), or unpleasant (n = 2), were easier to replace because all stimuli consisted of male faces (topics for the Neutral condition included facial hair and glasses). It was essential that the IAT Neutral condition was as neutral in its content as possible to be suitable as a control condition for fMRI analysis. Images that were rated as too pleasant (e.g., images of more handsome males) or unpleasant (e.g., images of unfriendly males) were therefore substituted with more average-looking males with neutral facial expressions (from the NimStim set of facial expressions).

Figure I5. Examples of a) Negative b) Positive and c) Neutral IAT stimuli. Pictures in a) were rated below 4 on the SAM Pleasure scale (except the Black face that was rated between 4 and 5), while pictures in b) were rated above 6 on the SAM Pleasure scale. Pictures in c) were rated as fairly neutral.

In terms of arousal ratings, none of the stimuli across different IAT conditions were regarded as particularly arousing. Although IAT Positive stimuli were statistically more arousing than IAT Negative stimuli, the actual differences in mean arousal values between these conditions were small, ranging from 5.02 to 5.38 (see Table I3). Moreover, none of the above-mentioned images in the Positive or Negative conditions that were rated as either too pleasant or unpleasant were associated with high levels of arousal. The finding of consistently neutral arousal values across experimental conditions was reassuring, especially in light of the fact that affective *arousal*, and not *valence*, has been demonstrated to be most strongly related to the degree of activation of the appetitive and defensive emotional motivation systems (Bradley, Codispoti, Cuthbert et al., 2001). In Bradley et al.'s experiment, pleasant as well as unpleasant pictures of low arousal were similar to neutral pictures in terms of the magnitude of physiological responses.

Given the data and results presented above, I thus felt confident that the selected IAT stimuli would have a negligible impact on any emotion elicitation paradigm during the imaging procedure.

Appendix J: IAT Result Sentences

Readability statistics were calculated for IAT result sentences in all stimulus conditions (i.e., Neutral, Positive, and Negative) to ensure that there were no discrepancies in the level of difficulty between conditions. Statistics calculated included the *Flesch Reading Ease score* and the *Flesch-Kincaid Grade Level score*. Both statistics rate text on a 1-100 point scale based on the same core values, namely word length and sentence length. They have, however, different weighting factors that lead to a rough inverse relationship between the two scores. Whereas it is easier to understand a document the higher the Flesch Reading Ease score is (standard documents are usually 60 to 70); the Flesch-Kincaid Grade Level formula translates the score to a U.S. school grade level. A score of 8.0, for example, means that the text would be easily comprehensible to an eighth grader.

To rule out the possibility of readability differences between experimental conditions, I compared the readability statistics of the Neutral, Positive, and Negative conditions using non-parametric Kruskal-Wallis tests. No differences emerged for the Flesch Reading Ease score ($p <$ .22) or the Flesch-Kincaid Grade Level score ($p <$ .22). The various stimulus conditions were thus of similar difficulty in terms of readability. Tables J1-J3 present the actual result sentences that were used after the completion of each IAT.

Table J1

fMRI Result Sentences Following Each IAT for the Neutral Stimulus Condition

	Result 1	**Result 2**	**Result 3**	**Stats**
Facial Hair IAT				
	Your data suggest little to no automatic preference for people with facial hair compared to people without facial hair	This is a regular/neutral response.	Extended Feedback: Generally, people with no preference for people with or without facial hair:	
Variant 1:			» Feel neutral about facial hair on other individuals. » Don't know exactly how often people shave. » Often won't recall whether colleagues regularly wear facial hair or not.	Count = 26 FRE = 51.6 FKG = 8.2
Variant 2:			» Feel neutral about cleanly shaved people. » Don't know a tremendous amount about different facial hair styles. » Often will not remember if someone was clean-shaven or not.	Count = 26 FRE = 54.8 FKG = 7.7
Glasses IAT				
	Your data suggest little to no automatic preference for people with glasses compared to people without glasses.	This is a regular/neutral response.	Extended Feedback: Generally, people with no preference for people with or without glasses:	
Variant 1:			» Feel neutral about people who regularly wear glasses. » Won't always notice if someone changes their glasses. » Often will not recall whether someone wears glasses or not.	Count = 26 FRE = 51.6 FKG = 8.2
Variant 2:			» Feel neutral about people who don't wear glasses. » Are unconcerned about friends wearing glasses or not. » Will fail to remember the specific colour of someone's glasses.	Count = 26 FRE = 54.8 FKG = 7.7

Note. Count = word count; FRE = Flesch Reading Ease score; FKG = Flesch-Kincaid Grade score

Table J2

fMRI Result Sentences Following Each IAT for the Positive Stimulus Condition

	Result 1	Result 2	Result 3	Stats
Sexuality IAT				
	Your data suggest absolutely no automatic preference for Straight compared to Gay people.	This is a low-prejudice/ positive response!	Extended Feedback: Generally, people with no automatic preference for Straight compared to Gay people:	
Variant 1:			» Treat all gay people with integrity and respect. » Believe that stereotyping a gay person is wrong. » Value internal qualities of a person rather than their sexual preference.	Count = 27 FRE = 47.3 FKG = 8.8
Variant 2:			» Try to be unbiased and positive toward gay people. » Will react positively to a homosexual person in company. » Are not judgmental towards people who are gay.	Count = 26 FRE = 51.6 FKG = 8.2
Religion IAT				
	Your data suggest absolutely no automatic preference for Other Religions compared to Judaism.	This is a low-prejudice/ positive response!	Extended Feedback: Generally, people with no automatic preference for Other Religions compared to Judaism:	
Variant 1:			» Do not have prejudiced ideas about Jewish individuals. » Have respect for other people and their spiritual values. » Believe that all people should have freedom of religion.	Count = 26 FRE = 51.6 FKG = 8.2
Variant 2:			» Do not have any disregard for Jewish people. » Judge people by their personal qualities rather than religion. » Believe strongly in the human rights of another person.	Count = 26 FRE = 51.6 FKG = 8.2

Note. Count = word count; FRE = Flesch Reading Ease score; FKG = Flesch-Kincaid Grade score

Table J3

fMRI Result Sentences Following Each IAT for the Negative Stimulus Condition

	Result 1	**Result 2**	**Result 3**	**Stats**
Race IAT				
	Your data suggest a significant automatic preference for White compared to Black people.	This is a high prejudice/ negative response!	Extended Feedback: Generally, people with an automatic preference for White compared to Black people:	
Variant 1:			» Are more afraid of Black people than White people. » Can't help having stereotypical thoughts about Black people. » Think Black people are generally less intelligent than White people.	Count = 27 FRE = 53.5 FKG = 8.0
Variant 2:			» Think Black people in authority are often corrupt. » Regard Black people as more lazy than White people. » Think White professionals are often more competent than Black professionals.	Count = 27 FRE = 53.5 FKG = 8.0
Disability IAT				
	Your data suggest a significant automatic preference for Abled compared to Disabled people.	This is a high prejudice/ negative response!	Extended Feedback: Generally, people with an automatic preference for Abled compared to Disabled people:	
Variant 1:			» Attach high value to things like IQ and physical appearance. » Think they are better than mentally disabled people. » Prefer spending time with abled rather than disabled people.	Count = 27 FRE = 53.5 FKG = 8.0
Variant 2:			» Are sometimes a bit awkward around disabled people. » Feel that disabled people use up lots of public resources. » Don't fully understand the emotional needs of disabled people.	Count = 27 FRE = 47.3 FKG = 8.8

Note. Count = word count; FRE = Flesch Reading Ease score; FKG = Flesch-Kincaid Grade score