CHAPTER 13

Why Symbolic Processing of Affect Can Disrupt Negative Affect: Social Cognitive and Affective Neuroscience Investigations

Matthew D. Lieberman

In the Highland Indian villages of Guatemala, miniature "worry" dolls approximately 1 inch in height and made from small bits of wood, cloth, and string are given from parent to child. According to legend, parents are meant to say the following along with the presentation of the gift: "If you have a problem, then share it with a worry doll. Before going to bed, tell one worry to each doll, then place them beneath your pillow. Whilst you sleep, the dolls will take your worries away!" It is unclear whether these dolls have actually been imbued with the power to whisk away worry, however there is a great deal of evidence to suggest that the process of sharing one's worry, of putting bad feelings into words, can diminish one's emotional distress, at least under certain circumstances. This chapter examines the neurocognitive mechanisms of *disruption effects*, the process by which putting feelings into words can disrupt the feelings being verbalized.

The notion that labeling emotional states can help to dampen down or regulate negative emotional states is hardly new. In commentary on some of the oldest Buddhist texts, it has been written that "The skillful use of labeling … introduces a healthy degree of inner detachment since the act of apostrophizing [i.e. speaking to] one's moods and emotions diminishes one's identification with them" (Analayo, 2003, p. 113). Similarly, a number of western thinkers have written about disruption effects prior to the twentieth century. The philosopher Benedict Spinoza suggested that "An emotion which is a passion, ceases to be a passion as soon as we form a clear and distinct idea thereof" (1675/2000, p. 291). In the *Principles of Psychology*, William James wrote that "The present conscious state when I say … 'I feel angry' is not the…direct state of anger … it is the state of *saying-I-feel-angry*. The act of naming them has momentarily detracted from their force" (1890, p. 190).

In modern psychology, emotions are often thought to be relatively uncontrollable with direct attempts at regulating one's own emotional state often backfiring (LeDoux, 1996; Wegner, Erber, & Zanakos, 1993; Wegner, Shortt, Blake, & Page, 1990). Nevertheless, the legacy of disruption affects lives in various forms of talk therapies. Talk therapies such as cognitive-behavioral therapy and psychoanalysis vary greatly in their approach and the putative mechanisms supporting successful outcomes; however, they all involve individuals putting feelings into words with the hopes of managing or transforming those feelings.

The insight that putting one's feelings into words can have mental and physical health benefits was captured experimentally in work on disclosure through expressive writing (for a review, *see* Lepore & Smyth, 2002). In the 1980s, Pennebaker began a program of research (Pennebaker & Beall, 1986; Pennebaker, 1997) in which participants were asked to write about past negative experiences on four successive

days, and these participants were found to have visited the doctor less often over the following half-year compared to those who wrote about trivial experiences. Although numerous studies have shown health benefits of expressive writing across numerous domains, including blood pressure (McGuire, Greenberg, & Gevirtz, 2005), chronic pain (Broderick, Junghaenel, & Schwartz, 2005), cancer-related symptons (Stanton et al., 2002), lung functioning (Smyth et al., 1999), liver functioning (Francis & Pennebaker, 1992), and immune function (Booth, Petrie, & Pennebaker, 1997), a number of other studies have shown that expressive writing leads to improvements in emotional well-being and mental health more generally (Hemenover, 2003; Park & Blumberg, 2002). It is unclear which aspects of the writing produce the physical and mental health benefits (for a review of different accounts, *see* Baikie & Wilhelm, 2005); however, it is clear that merely thinking about negative experiences without being required to organize those thoughts into words does not have the same benefits and can actually be quite detrimental to mental health (Lyubormirsky, Sousa, & Dickerhoof, 2006; Nolen-Hoeksema, 2000).

Intentional versus unintentional emotion regulation

Although the effects of expressive writing look like the results of emotion regulation processes, the expressive writing paradigm lacks certain indicators associated with emotion regulation. When one thinks of emotion regulation, one typically thinks of having a very overt intention to change one's emotional experience or at least the outward manifestations of that experience (Gross, 1998). One imagines "grinning and bearing it" when publicly receiving news that someone else received the promotion you were hoping for. Most would also expect that carrying out this intentional emotion regulation would feel effortful (Richards & Gross, 2000). It is unclear to what extent putting feelings into words, either during expressive writing or in other forms, constitutes an intentional or unintentional form of emotion regulation.

This blurred line between intentional and unintentional regulation is present in some of the earliest work on emotion regulation conducted by Lazarus and others. In these studies (Dandoy & Goldstein, 1990; Lazarus & Alfert, 1964; Lazarus, Opton, Nomikos, & Rankin, 1965; Speisman, Lazarus, Mordkoff, & Davison, 1964), subjects' physiological arousal was measured, typically while they watched disturbing films. By providing a verbal narrative explaining the content of the films in different ways, changes in the physiological responses were obtained. For example, telling subjects that the scene they were about to see was created by actors appearing to get injured and that the injuries were fake led to diminished skin conductance responses while subjects watched the scene, relative to subjects not so informed. The framing of the scene changed the appraisal of the scene's meaning and thus had apparent regulatory effects (i.e., diminished skin conductance responses), but it is unclear whether the subjects engaged in anything they would themselves call emotion regulation. Decades of work on placebo effects have a similar phenomenology associated with them (Benedetti, Mayberg, Wager, Stohler, & Zubieta, 2005), such that a belief or appraisal that a pill will prevent pain actually leads to diminished experiences of pain, despite the pill having no active ingredients. More recent fMRI work (Ochsner, this volume) has put this reframing or *reappraisal* process in the hands of subjects and thus made the process fully overt, asking subjects to understand aversive stimuli in ways that make them less aversive.

The expressive writing studies (Pennebaker et al., 1997) and appraisal studies (Lazarus & Alfert, 1964) suggest that verbal processing of emotional content and explicit changes to the framing of emotional content can serve to regulate emotional responses, even when there is no obvious regulatory intent. Nevertheless, these paradigms could both produce spontaneous intentions to regulate one's emotions, and this could be serving as an unmeasured, but mediating, mechanism. Two other lines of research suggest that intention to regulate one's affect is not, in fact, necessary for the disruption of

affect to occur for a broader review of unintentional emotion regulation, see Berkman and Lieberman (2009).

For example, Wilson and Schooler (1991; Wilson et al., 1993) conducted a series of studies demonstrating that reflecting upon and writing about one's own affective state disrupted the impact that their affective states would otherwise have had on their decision making. Critically, in these studies, the task was not focused on emotion regulation at all but instead was focused on merely making good decisions by consulting one's own affective response as a guide. In one study, individuals were asked to choose between a number of works of art and were ultimately able to take one art print home with them. Some individuals were also asked to reflect on their feelings about each of the prints *before* announcing their rating. Surprisingly, individuals who reflected on their feelings before choosing were more likely to choose an art print that they themselves would later regret choosing than individuals who did not reflect on their feelings. The authors suggested that some aspects of feeling states are more verbalizable than others, and when making a decision, we weight verbal information in our minds more heavily than nonverbal feelings. Thus, if good decisions are driven by feelings that cannot be easily verbalized, relying on that which can be verbalized will produce suboptimal decisions. It is also possible, however, that verbalizing one's feelings temporarily altered the feeling states themselves by dampening them. Behavioral data alone cannot easily tease these two interpretations apart (i.e., overemphasizing verbal information vs. dampening of affect) and this was actually one of the original incentives for using fMRI to examine this issue, as it may be better suited for teasing apart these interpretations.

Another study by Greenberg, Wortman, and Stone (1996) more directly addresses the issue of whether regulatory intent is critical for the benefits of putting feelings into words. In this study, an expressive writing paradigm similar to Pennebaker's was used except that an additional condition was included. Individuals in this condition were asked to write about a trauma, but one that was imagined rather than real. Despite the imaginary nature of the traumas written about, these individuals showed benefits of expressive writing similar to those seen in previous studies. It is difficult to argue that these benefits derived from any overt attempts at emotion regulation. Instead, merely putting feelings into words—albeit imagined feelings—produced disruption-like effects.

It is important to note here that I am not suggesting that intentional emotion regulation is reducible to putting feelings into words. The understanding that people have of themselves and of those around them guide their emotional lives, and thus new understandings reached through introspection, disclosure, and reappraisal undoubtedly have the power to transform one's emotional responses. I am simply suggesting that *some of the benefits* derived from these therapeutic techniques may result from neurocognitive consequences of merely putting feelings into words. And if this is the case, these benefits could be put to good use therapeutically, even in cases for which an individual is unwilling or unable to engage in emotion regulation.

RVLPFC AS A CANDIDATE MECHANISM

The rest of this chapter is devoted to exploring one possible neurocognitive mechanism by which putting feelings into words could disrupt basic negative affect processes, thereby improving one's affective state. Disruption theory posits that right ventrolateral prefrontal cortex (RVLPFC; *see* highlighted area in Fig. 13–1d) plays a central role in the disruption effects. RVLPFC long been associated with inhibitory processes and more recently it has been identified in studies examining the symbolic processing of affect. With both of these functions associated with RVLPFC activity, RVLPFC emerges as an ideal candidate for disruption effects, as these effects appear to involve symbolic processing of affect, which leads to the inhibition of affective processes. Before turning to the evidence that experimentally combines these functions in RVLPFC, I first review the evidence that links RVLPFC separately to inhibition and to symbolic processing of affect.

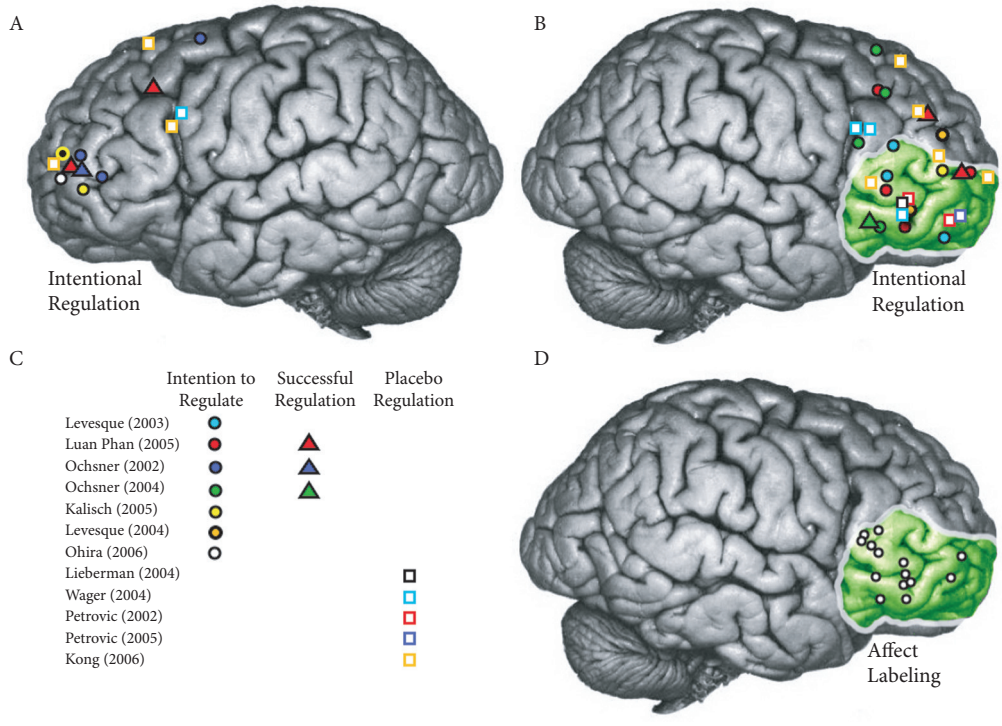

Fig. 13–1 Right ventrolateral prefrontal activity (RVLPFC; highlighted area) in affect labeling and emotion regulation studies. (A) Left lateral and (B) Right lateral activations in studies of emotion regulation and placebo effects. (C) Legend for emotion regulation and placebo effects (D) RVLPFC activations in affect labeling studies.

RVLPFC and Inhibition

Although there is ongoing debate about the full set of neural regions involved in inhibitory processes, RVLPFC would certainly be included in anyone's candidate set. More than a dozen neuro-imaging studies of the Go-NoGo, Flanker, and Stroop tasks have identified RVLPFC activations associated with trying to inhibit a prepotent motor response or trying to ignore task-irrelevant information that would lead to an incorrect response (Asahi, Okamoto, Okada, Yamawaki, & Tokota, 2004; Blasi et al., 2006; Garavan, Ross, & Stein, 1999; Horn, Dolan, Elliott, Deakin, & Woodruff, 2003; Kawashima, 1996; Konishi, 1999; Liddle, Kiehl, & Smith, 2001; Matthews, Simmons, Arce, & Paulus, 2005; Rubia, Smith, Brammer, & Taylor, 2003; Hazeltine, Poldrack, & Gabrieli, 2000; Hazeltine, Bunge, Scanlon, & Gabrieli, 2003; Fan, Flombaum, McCandiss, Thomas, & Posner, 2003; Kemmotsu, Villalobos, Gaffrey, Courchesne, & Muller, 2005; Leung, Skudlarski, Gatenby, Peterson, & Gore, 2000). In addition, these tasks have found that RVLPFC activity is associated with faster reaction times on inhibition trials (Garavan et al., 1999), that RVLPFC activity is greater for successful inhibition trials than unsuccessful inhibition trials (Rubia et al., 2003), and that RVLPFC activity is greater for harder inhibition trials than easy inhibition trials (Matthews et al., 2005). Children with attention deficit hyperactivity disorder (ADHD) show impaired behavioral performance on motor inhibition tasks and also evidence less RVLPFC activity during inhibition tasks than controls (Durston, Mulder, Casey, Ziermans, & van Engeland, 2006; Rubia et al., 1999). One study that observed better motor inhibition in an ADHD sample after neurofeedback training

also observed an increase in RVLPFC activity, relative to a sample that did not receive this training (Beauregard & Levesque, 2006). Studies of permanent lesions (Aron, Fletcher, Bullmore, Sahakian, & Robbins, 2003) and temporary lesions to RVLPFC induced by transcranial magnetic stimulation (Chambers et al, 2006) have also found impaired motor inhibition. Finally, pharmacological studies in which participants receive serotonergic agonists, associated with enhanced self-control and diminished impulsivity, observed greater activity in RVLPFC during motor inhibition trials (Anderson et al., 2002; Del Ben et al., 2005; *see also* Rubia et al., 2005, but cf. Vollm et al., 2006).

A fascinating study by Goel and Dolan (2003) suggests that RVLPFC may also be involved in nonmotoric forms of inhibition such as the inhibition of belief. In this study, participants assessed the validity of syllogisms (i.e., Does the conclusion logically follow from the premises?) that were either sound (premises were true) or unsound (one premise was false). Participants had difficulty accurately identifying a valid syllogism as valid if it was unsound and therefore not true. For example, given the premises "All addictive things are expensive" and "Some cigarettes are inexpensive," it is valid to conclude that "Some cigarettes are not addictive" although the first premise and conclusion are false. RVLPFC was the only region of the brain that was more active when participants overcame their belief-bias and indicated that this kind of syllogism was valid. A number of studies on active deception have also suggested a role for RVLPFC in the inhibition of belief (Abe et al., 2006; Spence et al., 2001; Luan Phan et al., 2005; Nunez, Casey, Egner, Hare, & Hirsch, 2005). Across these studies, when individuals were required to inhibit what they knew to be true to say something false, RVLPFC was recruited.

RVLPFC and Symbolic Processing of Affect

There have been many fewer studies examining symbolic processing of affect (SPA) than inhibitory processes, but the percentage of SPA studies implicating RVLPFC is at least as high as that seen in the inhibition literature. SPA refers, roughly, to the explicit linguistic/propositional processing of one's own affect ("I feel sad"), the affect of others ("She looks frightened"), evaluatively valenced categories ("Terrorists are bad"), or the value of response options ("I will lose money if I keep my money in betamax stock"). Across a variety of studies, RVLPFC tends to be more active during SPA than non-SPA, particularly in the case of negatively valenced SPA.

For example, Cunningham and colleagues (Cunningham, Johnson, Gatenby, Gore, & Banaji, 2003) presented participants with famous names like Bill Cosby and Adolph Hitler, who are generally viewed either positively or negatively. On some trials, participants were asked to decide whether the target was alive or dead but on other trials were asked if the target was good or bad. Thus, on all trials, implicit affective responses to the targets should be expected, but explicit SPA should only occur when the targets are evaluated as good or bad. Cunningham et al. (2003) observed that RVLPFC along with medial prefrontal cortex (mPFC) were more active during good/bad judgments than during alive/dead judgments, suggesting that these regions are involved in SPA. They also found that RVLPFC was the region of the brain that was most active during bad judgments relative to good judgments, suggesting a possible selective role in negative SPA.

A number of studies that have focused on explicit judgments about the emotional aspects of pictures (Gorno-Tempini et al., 2001; Gur et al., 2002; Nakamura et al., 1999; Narumoto et al., 2000; Royet, Plailly, Delon-Martin, Kareken, & Segebarth, 2003) and voices (Wildgruber et al., 2004, 2005) demonstrated greater RVLPFC activations to emotional than nonemotional judgments. A study that specifically compared negative emotion judgments to neutral and positive judgments observed greater RVLPFC to negative emotion judgments (Dolcos, LaBar, & Cabeza, 2004), similarly to Cunningham et al. (2003). In addition, multiple studies have observed that reading negatively valenced words is associated with greater RVLPFC than reading neutral or positive words (Cunningham, Espinet,

DeYoung, & Zelazo, 2006; Cunningham, Raye, & Johnson, 2004; Kuchinke et al., 2005).

Nomura et al. (2003; *see also* Shaw et al., 2005) compared difficult emotion judgments to easy emotion judgments. Presumably, the difficult judgments required more top-down elaboration of the emotional qualities of the stimulus than the easy judgments and thus would involve more SPA. In this study, participants judged the emotional expression or the gender of target faces. For half of the trials, the critical dimension was ambiguous (e.g., half of the gender trials had faces that were ambiguous with respect to gender). Nomura et al. (2003) found that RVLPFC and the dorsal anterior cingulate cortex (dACC) were the only regions of the brain that were more active during ambiguous trials than unambiguous trials. Importantly, however, the effect in RVLPFC was driven entirely by its response to ambiguous emotion trials, whereas the dACC was equally responsive to both kinds of ambiguity. Thus, one reasonable interpretation of these results is that RVLPFC was recruited on ambiguous emotion trials as participants engaged in explicit hypothesis testing about the emotional expression, which would be consistent with its putative role in SPA.

RVLPFC ANATOMICAL PROJECTIONS TO LIMBIC REGIONS

The preceding sections set up the possibility that SPA in RVLPFC could inhibit activity in limbic regions such as the amygdala, insula, and ACC associated with affective experience. It is important to establish that such a claim is neuro-anatomically plausible. That is, does RVLPFC have the right kinds of neuro-anatomical connections to these other regions to produce these regulatory effects? For the connections to the insula and ACC, the answer is a resounding yes. RVLPFC has strong bidirectional connections with both of these regions (Augustine, 1996; Vogt & Pandya, 1987).

The neuro-anatomical connections from RVLPFC to the amygdala are more complex. On the one hand, there are direct projections from RVLPFC to the amygdala. Carmichael and Price (1995; *see also* Ghashghaei & Barbas, 2002; McDonald, Mascagni, & Guo, 1996) made anterograde tracer injections into area 12l (the region in the rhesus monkey homologous to Brodmann's area 47 in humans) and found evidence of projections from area 12l to the basolateral nucleus of the amygdala (BLA). However, these projections are not particularly dense, calling into question whether these direct projections are sufficient to allow RVLPFC to regulate amygdala responses. As suggested by Phelps, Delgado, Nearing, and LeDoux (2004), RVLPFC could also have its effect on the amygdala indirectly by way of projections from RVLPFC to mPFC, which in turn has dense projections to the amygdala (Carmichael & Price, 1995) and is known to regulate the amygdala in studies of extinction (Phelps et al., 2004; Quirk, Likhtik, Pelletier, & Pare, 2003).

RVLPFC DIMINISHES NEURAL AND SUBJECTIVE NEGATIVE AFFECT

This section reviews research that suggests that RVLPFC not only inhibits motor and cognitive responses but also inhibits negative affective responses both in terms of subjective reports of negative affect and in terms of activity in limbic regions associated with negative affect and distress. In light of the previous sections that establish a major role for RVLPFC in *(1)* inhibitory processes; *(2)* the symbolic processing of negative affect; and *(3)* possessing neuro-anatomical connections to limbic regions, it is perhaps not a giant leap to suggest that RVLPFC may contribute to the inhibition of motoric, cognitive, and affective responses. Nevertheless, establishing this relationship will serve as a critical stepping stone to full-blown disruption effects reviewed in the next section.

RVLPFC is one of the regions that has been associated with increased pain analgesia (Petrovic, Kalso, Petersson, & Ingvar, 2002). More recently, a number of studies have observed that placebo effects appear to be mediated by RVLPFC, along with rostral anterior cingulate cortex (rACC). In one study, we (Lieberman et al., 2004) examined a group of patients with irritable bowel syndrome (IBS), a chronic pain condition associated with heightened pain

sensitivity in the limbic system (Naliboff et al., 2006). The IBS patients were scanned prior to and then again after receiving 3 weeks of sham treatment with placebos for their pain. During each scanning session, patients received painful rectal stimulation, simulating the symptoms of IBS and generating a measure of current neural responses to this stimulation. We found that to the extent that participants reported improvements in their pain symptoms at the end of the placebo regimen, compared to before the regimen began, they also showed increased activity in RVLPFC ($r = 0.71$) and decreased dACC activity from the first scanning session to the second. Multiple other studies have also observed within session placebo effects associated with increased RVLPFC activity and decreased limbic activity in the domains of physical pain (Petrovic et al., 2002; Wager et al., 2004) and anxiety (Petrovic et al., 2005).

We have also examined the role of RVLPFC in the regulation of "social pain" or the distress associated with social rejection (Eisenberger, this volume; Eisenberger & Lieberman, 2004). In one study (Eisenberger, Lieberman, & Williams, 2003), participants ostensibly played a game of Internet "catch" with two other players, who were actually computer simulations. Part of the way through the game, the other players stopped throwing the ball to the participant and thus excluded the participant for the rest of the game. Numerous behavioral studies have shown that this exclusion manipulation causes considerable distress in participants, even when they know the other players are just computer simulations (Williams, in press). Our participants also reported being distressed in response to being excluded and showed a pattern of neural activity consistent with the experience of visceral pain (*see also* Eisenberger, Way, Taylor, Welch, & Lieberman, 2007). Most relevant here is that participants produced increased activity in dACC to the extent that they felt; however, to the extent that RVLPFC was active, participants reported feeling less distressed by the episode of exclusion. Moreover, activity in RVLPFC was negatively correlated with dACC activity, and changes in dACC activity mediated the relationship between RVLPFC and distress. In other words, it appears that increased RVLPFC activity may have helped to downregulate dACC responses, which in turn were associated with reduced distress.

In contrast to the social and physical pain studies, fMRI studies of reappraisal explicitly instruct subjects to engage in emotion regulation. Nearly all of the fMRI studies of reappraisal have observed activity in or near RVLPFC along with other prefrontal regions (*see* Fig. 13–1a & 13–1b: Beauregard, Levesque, & Bourgouin, 2001; Kalisch et al., 2005; Levesque et al., 2003; Luan Phan et al., 2005; Ochsner et al., 2004; Schaefer,et al., 2003; cf. Ochsner, Bunge, Gross, & Gabrieli, 2002).

A handful of other studies have implicated RVLPFC in the regulation of emotional behaviors. These studies may be something of a blend between the motor inhibition and emotion regulation paradigms, supporting the notion that RVLPFC is involved in a continuum of regulatory effects. In one study (Small, Zatorre, Dagher, Evans, & Jones-Gotman, 2001), participants were required to eat a piece of chocolate during each of a series of PET scans. After each scan, participants indicated how much they liked eating the chocolate and how much they wanted to have another piece. Predictably, in early scans, participants liked the chocolate and wanted more; however, by the second half of the study, the participants did not like the chocolate anymore and did not want to eat another piece. Activity in RVLPFC was strongly associated with self-reports of not wanting to eat anymore chocolate despite being asked by the experimenter to continue eating it, suggesting that RVLPFC may have been involved in suppressing the desire to reject the chocolate to comply with the requirements of the study (i.e., eating the unwanted chocolate). Note that although not framed as such in this study, the results may have implications for future work on the neural correlates of compliance and conformity.

In another recent study (Tabibnia, Satpute, & Lieberman, 2008b), we examined how individuals overcome the slight of insulting unfair offers in a financial bargaining game to accept financially advantageous offers. Participants

played the "responder" role in several one-shot versions of the ultimatum game. In this game, the "proposer" is asked to split a sum of money between him/herself and the responder. Thus, if the proposer has a $10 stake to split, she may propose an even split of $5 and $5 or, perhaps, a more unfair split of $8 for herself and $2 for the responder. The responder then decides whether or not to accept the offer. If the responder accepts, then both the proposer and responder get exactly what the proposer proposed. However, if the responder rejects the proposal, neither participant receives anything. Either way, there is no additional bargaining after the responder responds.

An earlier fMRI study of the ultimatum game (Sanfey, Rilling, Aronson, Nystrom, & Cohen, 2003) compared the neural responses to fair ($5 out of $10) and unfair ($1 out of $10) offers. The main finding was that unfair offers were associated with increased activity in the anterior insula, a region that has previously been associated with disgust responses. In our study (Tabibnia et al., 2008b), we also included offers that were unfair and yet still financially desirable to undergraduate participants. In the study by Sanfey et al., both kinds of offers presented little conflict as the $5 offers were both fair and desirable, financially, whereas the $1 (and $2) offers were unfair and not that desirable, financially. To create this conflict between fairness and financial desirability, we included offers such as $5 out of $23, which were both insulting and yet also financially desirable. What we found across a number of different analyses is that the tendency to reject unfair but financially desirable offers was associated with activity in the anterior insula, consistent with the results from Sanfey et al. However, the tendency to accept the unfair but financially desirable offers was associated with activity in RVLPFC. Moreover, greater RVLPFC activity on these trials was associated with diminished anterior insula activity, and changes in anterior insula activity mediated the relationship between RVLPFC activity and the tendency to accept unfair offers. These results are consistent with the idea that RVLPFC is involved in dampening the limbic response to the insulting offer, allowing the individual to "swallow one's pride" and accept the unfair offer.

Symbolic processing of affect disrupts affect via RVLPFC

I have established that RVLPFC activity is associated with the inhibition of motor, cognitive, and emotional responses. Additionally, RVLPFC is active in various forms of SPA, particularly negatively valenced SPA (SPA_{Neg}). If SPA_{Neg} activates RVLPFC and activity in RVLPFC is associated with the inhibition of emotional responses, then it seems plausible that SPA_{Neg} would be associated with the inhibition of emotional responses and that activity in RVLPFC would be largely responsible for this effect.

Prior to the studies that directly linked SPA with the downregulation of affect, there were also a handful of studies suggestive of this link without overtly assessing it. Hornak, Rolls, and Wade (1996) tested a sample of patients with ventral prefrontal damage and found that these patients were impaired at explicitly recognizing emotional face expressions and voice tones. Of the 11 patients in the sample, 9 had right or bilateral ventral damage, and 8 of these were impaired on one or both SPA tests. Of the 2 "left-only" ventral prefrontal patients, one performed well above the mean of the nonventral controls. Additionally, the extent of impairment in SPA tasks was correlated with disinhibition of emotional behavior, suggesting that impaired ability to engage in SPA is associated with more emotional behavior and that this association may be related to ventral prefrontal impairment.

Hariri, Bookheimer, and Mazziotta (2000) produced the first evidence of the complete pathway from SPA to RVLPFC activity to reduced amygdala activity. In their study, participants judged the emotional identity of a target's facial expression, however, the trials varied with respect to whether symbolic processing was required to make the judgment. In the SPA condition ("affect label"; *see* Fig. 13–2a), a target face was presented at the top of the screen along with two emotion words (e.g., "angry," "surprised") at the bottom of the display, and participants had to choose which of

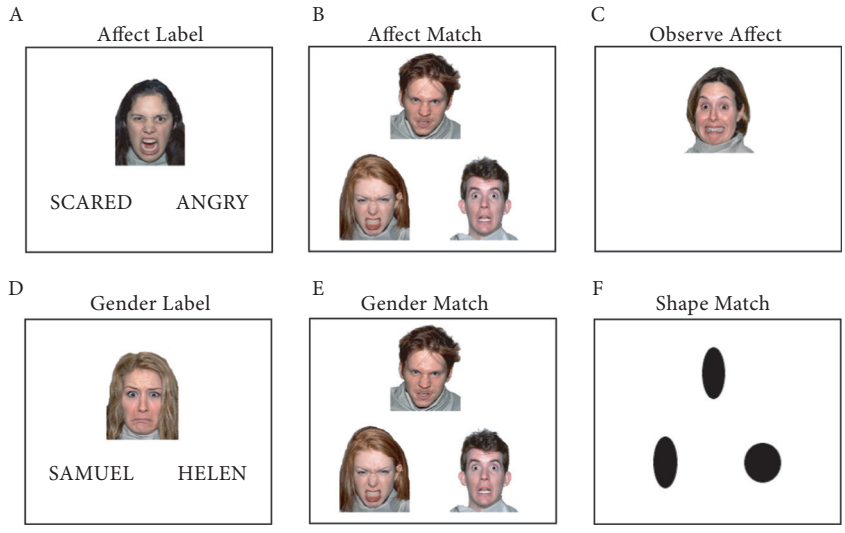

Fig. 13–2 Sample trials from an affect labeling study (Lieberman et al., 2007).

words best described the target's emotion. In the non-SPA condition ("affect match"; *see* Fig. 13–2b), a target face was presented at the top of the screen along with two other emotional faces at the bottom of the display and participants had to choose which of these were showing the same emotion as the target face. According to Hariri et al., in the non-SPA condition participants could "match the faces based on perceptual characteristics, such as wide eyes, furrowed brow or clenched teeth, but need not judge or interpret the information" (p. 44). Indeed, when viewing these stimuli, there is a strong sense of "pop-out" in the non-SPA stimuli in which the faces that match seem to automatically pop-out together.

In the non-SPA condition, there was significant amygdala activity relative to a shape-matching control condition ("shape match"; *see* Fig. 13–2f); however, there was no amygdala activity observed during the SPA condition. Instead, SPA was associated with activity in RVLPFC and the fusiform "face" area, the latter presumably indicating that the target face was still being attended to in the SPA condition. In the direct comparison of SPA and non-SPA trials, greater RVLPFC and diminished amygdala activity was observed during the SPA trials. Thus, two different forms of emotional processing—one symbolic and one non-symbolic—appear to be routed through distinct neural systems. Given that the amygdala has been shown in multiple studies of affective processing to be activated by conditions that would allow only automatic processing (i.e., subliminal presentations and binocular rivalry studies; Morris et al., 1998; Whalen et al., 1998; Pasley, Mayes, & Schultz, 2004; Villeumier, Armony, Driver, & Dolan, 2001), it is quite surprising to see the amygdala not responding under conditions that would allow both automatic and controlled processing.

In a follow-up study, we (Lieberman, Hariri, Jarcho, Eisenberger, & Bookheimer, 2005) compared SPA and non-SPA processing in the context of race. Rather than using different facial expressions of emotions, we used all neutral expression faces that varied by race. In the United States, the stereotypes of Blacks are evaluatively negative, particularly when assessed implicitly (Devine, 1989). Indeed, even U.S. Blacks have more negative implicit stereotypes of Blacks than of Whites (Nosek, Banaji, & Greenwald, 2002; Livingston & Brewer, 2002). Consistent with these behavioral findings, a number of neuro-imaging studies have observed greater amygdala activity to Black faces than to White faces, at least to the extent that participants

possessed strong anti-Black implicit stereotypes (Phelps et al., 2004; Cunningham, Johnson, Raye, Gatenby, Gore, & Banaji, 2004). We reasoned that because a neutrally expressive Black face produces a similar amygdala response as a negatively expressive White face, engaging in SPA by labeling the race of Black target faces might disrupt this race-related amygdala activity in much the same way that affect labeling disrupts the amygdala response to negatively expressive faces. It is worth noting that another reasonable hypothesis is that race labels would focus attention onto the negative stereotyped aspect of the targets (i.e., race) rather than on other more neutral or positive aspects (i.e., gender) and would therefore produce greater activity in the amygdala.

As in other race fMRI studies, we observed greater amygdala to Black faces than White faces when participants performed a "race-match" task (visually analogous to the trial shown in Fig. 13–2b) that did not require SPA. In fact, we observed this separately for both our White and Black participants. That is, Black participants produced greater amygdala activity to Black faces than White faces, consistent with the previous behavioral findings of Blacks displaying negative implicit stereotypes towards Blacks (Nosek et al., 2002; Livingston & Brewer, 2002).

In contrast to the non-SPA condition, when participants performed the "race-label" task (analogous to Fig. 13–2a), there was no differential amygdala activity to Black and White faces, and the amygdala responses to Black faces diminished compared to the amygdala response during race matching of Black faces and even compared to the control task that did not involve faces at all. As predicted, there was greater RVLPFC activity during race labeling of Black faces (SPA_{Neg}) but not during the race labeling of White faces (SPA_{Pos}). Additionally, there was a strong negative correlation between RVLPFC and amygdala activity during race labeling of Black faces such that the individuals who activated RVLPFC the most during these blocks also tended to activate the amygdala the least. Finally, all of these effects were evident for both the Black and White participants.

Disruption effects redux

The advantage of the affect labeling paradigm over previous SPA studies is that during both matching and labeling conditions, attention is focused on the emotional aspects of the stimulus, with only the need to engage in SPA varying across the conditions. Affect labeling requires SPA, whereas affect matching does not, although affect matching does not prevent spontaneous SPA. Additionally, by using verbal labels that appear in different positions across trials, participants cannot learn a stimulus response mapping between, say, perceptual cues of fear and a right button press. Participants need to read the labels on each trial to see which options are available.

Despite these advantages, there are some inferential limitations present in the original formulation of the affect labeling paradigm. Although the comparison of the affect label to the affect match conditions represents a comparison of SPA and non-SPA, this distinction is confounded with other differences between the conditions. First and foremost, affect match trials present three faces, of which at least two are posing negative emotional expressions on most trials. In contrast, the affect label trials never present more than a single negatively expressive face. Thus, one could argue that greater amygdala activity is present in the affect matching condition because there are more amygdala activating stimuli present on those trials. This argument is not entirely satisfactory given that a single negatively expressive face, even presented subliminally, is usually sufficient to produce amygdala activity (Morris et al., 1998; Whalen et al., 1998), whereas neither of the two affect labeling studies reported the presence of amygdala activity during the affect labeling condition.

Another possibility is that affect labeling is not really affecting amygdala activity, but rather, affect matching leads to hyper-amygdala responses and thus the difference between the two conditions emerges. This criticism does not address the issue of why there has been no amygdala activity observed during the affect labeling condition, but it does raise the important issue

that affect matching has different task requirements than tasks that typically provoke amygdala activity such as passive observation of faces or making gender judgments of faces. It is unknown how much the difference between the labeling and matching conditions results from each of these factors because a passive observation condition has not been included.

The last criticism of the paradigm acknowledges that the labeling condition is indeed modulating amygdala activity but takes issue with the source of this modulation. Although we have characterized the affect labeling task in terms of SPA and non-SPA, one could just as easily label them as cognitive and perceptual processes more generally without making any claims about the affective component of these tasks. In other words, perhaps any kind of cognitive or verbal labeling process will diminish the amygdala response to these emotional stimuli.

To address all of these concerns, we ran a modified version of the affect labeling task that included a number of control conditions (Lieberman et al., 2007). All of the conditions of this study are shown in Figure 13–2. We included a passive observation condition (Fig. 13–2c) during which subjects were presented with a single negative emotional target face on each trial and simply attended to the face. This condition was used to construct regions of interest (ROIs) in the amygdala, which could then be compared across all conditions to examine the modulatory effect of other forms of stimulus processing. Then, in addition to the standard affect label and affect match conditions, we included gender label and gender match conditions (Fig. 13–2d & 13–2e). The comparison of affect label and gender label is the most critical comparison as both conditions present only a single target face and both involve labeling—albeit different kinds of labeling (affective vs. non-affective).

As can be seen in Figure 13–3, affect match, gender match, and gender label each produced amygdala activity that was statistically equivalent to that produced during the passive observation of emotional faces ("observe"). Only affect labeling produced significantly less amygdala than the observe condition. Affect labeling also produced less amygdala activity than gender labeling or affect matching, indicating that this effect really resulted from SPA rather than the number of faces on each trial or cognitive processes more generally. Incidentally, in whole-brain analyses, a number of limbic and paralimbic structures were also less active during affect labeling than gender labeling, including dorsal ACC, subgenual ACC, posterior insula, and ventromedial PFC.

In contrast, only a single region of the brain, RVLPFC, was more active during affect labeling

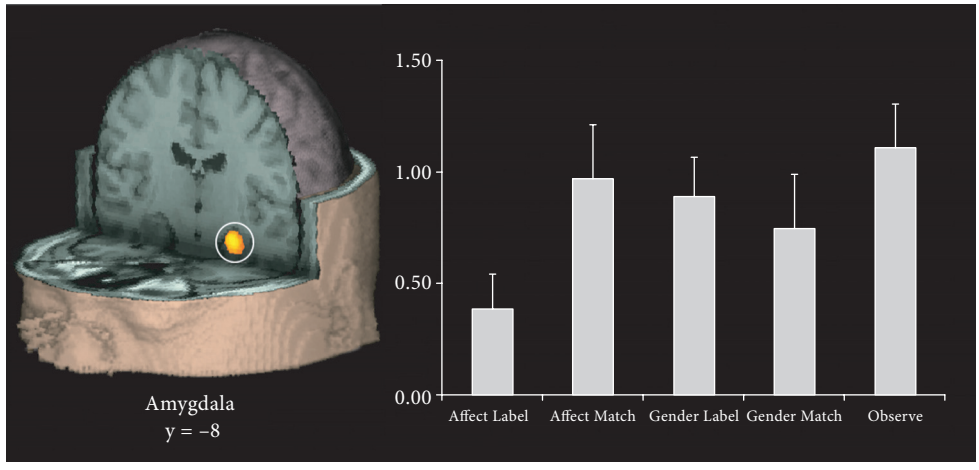

Fig. 13–3 Amygdala response under various processing conditions. Only affect labeling produced a lower level of amygdala activity than simply observing a negative emotional face.

than gender labeling. In addition, after running a correlational analysis using the amygdala cluster from the comparison of affect and gender labeling as a seed, we found that RVLPFC was one of only two regions that had a negatively correlated pattern of activity during this comparison. In other words, if one wanted to know which subjects produced the least amygdala activity during affect labeling, relative to gender labeling, finding the subjects who had the most activity in RVLPFC would be the way to do this. Interestingly, mPFC in BA10 was the only other region of the brain to show this pattern. This is interesting because mPFC has been identified as a possible mediator of RVLPFC effects on the amygdala. Additionally, mPFC is critical to extinction processes and the regulation of the amygdala in this context (Phelps et al., 2004; Quirk et al., 2003) and has been associated with reflective emotional processes (Lane et al., 1997; Taylor, Phan, Decker, & Liberzon, 2003). In running a mediational analysis, we found support for the RVLPFC→mPFC→amygdala pathway effect such that the relationship between RVLPFC and the amygdala during affect labeling was significantly mediated by mPFC activity.

In a psychophysiological follow-up, we found similar results for skin conductance, paralleling the amygdala findings in the fMRI research. In this study (Crockett, Lieberman, & Tabibnia, unpublished manuscript), subjects performed the affect label, affect match, gender label, and shape match tasks while skin conductance responses (SCR) were measured. Across the entire sample, affect labeling was associated with smaller SCRs than affect matching and equivalent SCRs to the shape-matching control task. Gender labeling produced SCRs between the levels observed for affect labeling and affect matching but was not significantly different from either. One reason these latter effects may not have been significant is that a number of subjects did not show reliable SCRs in any of the conditions, which dampened the statistical power of the entire sample. This may have occurred because face stimuli are not as emotionally provocative as other stimuli known to produce strong SCRs (Britton, Taylor, Sudheimer, & Liberzon, 2006), such as the images from the International Affect Picture System (IAPS; Lang, Bradley, & Cuthbert, 1999). When we separated the sample into high and low neurotics, a clearer picture emerged. Non-neurotics, who tend to be less reactive to negatively valenced stimuli, showed no reliable SCR differences across any of the conditions. Those high in neuroticism, however, produced strong SCR responses to affect match and gender label trials and much weaker SCR responses to affect label and shape match trials. Thus, for those that were showing SCR responses at all to the emotional stimuli, the disruption hypotheses were fully supported.

Affect labeling and behavioral inhibition

RVLPFC activity is associated with reduced activity in limbic regions, such as the amygdala and dACC, and SPA is associated both with increased RVLPFC activity and decreased limbic activity. One of the core reasons for pursuing this line of work is the established role of RVLPFC in motor and behavioral inhibition. In light of these various effects, it is reasonable to ask whether SPA, which activates RVLPFC, also has inhibitory effects on behavior. Perhaps RVLPFC produces various forms of inhibition simultaneously (although past studies have typically looked at motor, cognitive, or affective inhibition alone), and perhaps SPA sets the various forms of inhibition in motion. This would certainly be consistent with claims of Goethe, Emerson, Dewey, Arendt and others that thought paralyzes action. In a recent study, Robinson and Wiklowski (2006) found behavioral evidence indicating that SPA_{Neg} leads to motor inhibition, observing that reading negatively valenced primes, but not neutral or positive prime words, led to longer reaction times on a simple motor response task.

We conducted an fMRI study (Lieberman, Eisenberger, & Crockett, unpublished manuscript) to examine the effects of priming a negative stereotype on walking speed. We adapted the classic "automatic behavior" study (Bargh, Chen, & Burrows, 1996) in which priming the

"elderly" stereotype leads to slower walking, for use in the scanner environment. We reasoned that reading sentences related to the negative valenced stereotype of the "elderly" constitutes a form of SPA_{Neg} just as labeling the race of Black targets did in our previous study (Lieberman et al., 2005). If true, this would be expected to activate RVLPFC and diminish the activation in limbic structures and possibly inhibit motor processes as well, which could promote slower walking.

This is exactly what we found. After being primed with sentences related to the elderly stereotype in the scanner, participants walked more slowly than they did before scanning. Although part of this effect was no doubt the result of the general sluggishness felt after scanning, we were interested in how neural activity during the sentence priming related to the changes in walking speed from pre- to post-scanning. We found that RVLPFC was the only region of the brain for which greater activity during the priming of the elderly stereotype was associated with more slowing from pre- to post-scan walking measurements. As in our previous studies, we also observed greater increases in RVLPFC associated with reductions in limbic areas, including the amygala and dACC. However, greater activity in RVLPFC was also associated with less activity in the cerebellar vermis, a region that has been associated with motor processes related to walking and lower limb control (Jahn et al., 2004; Martin, 1996). Moreover, during the presentation of sentences related to the elderly, compared to control sentences, this same region of cerebellum was less active. Thus, in this study, SPA_{Neg} not only activated RVLPFC and attenuated limbic responses but also attenuated activity in a region linked to motor preparation and to walking behavior, suggesting that SPA_{Neg} may, in fact, produce motor inhibition as well as emotion regulation. It should also be noted that the RVLPFC-limbic effects occurred in this study despite any plausible impetus for subjects to intentionally engage in emotion regulation. Consequently, it appears that the desire to regulate one's emotional responses may not be necessary to receive the regulatory benefits of activating RVLPFC, consistent with previous research on the benefits of writing about imaginary traumas (Greenberg et al., 1996).

Clinical applications

Given that SPA appears to regulate limbic responses without the intention to do so, this would provide a mechanism by which putting feelings into words would have benefits for regulating emotional distress and for mental health more generally. In an initial attempt to bridge between disruption studies and clinical therapy, we have conducted a series of studies that integrate a SPA manipulation into an analogue of exposure therapy.

In one study, Tabibnia, Lieberman, and Craske (2008a) presented participants with a number of different high-arousal negative images from the IAPS (Langet al., 1999) on Day 1 while SCR was measured. Each of the pictures was presented a total of six times throughout the session to mimic the repeated exposure involved in exposure therapy (Foa & Kozak, 1986). Some of the pictures were presented alone on each trial, whereas others were presented and then followed by either a neutral or negatively valenced word on each trial. Once a picture was presented alone, with a negative word, or with a neutral word, the picture was presented the same way for all the trials. However, the specific words used varied with each presentation, such that a picture presented with negative words would be presented with six different negative words across the six presentations, thus preventing strong associations to a particular word. Exposure therapy is based on the premise that allowing individuals to fully experience an emotional response to a feared stimulus on multiple occasions will allow that emotional response to subside over time. In light of this, the temporal placement of the affect labels was deemed critical. We presented the words 3.5 seconds after the pictures to allow a full physiological response to emerge. Because disruption theory posits that the labels can reduce these responses, simultaneous presentation of pictures and words might actually prevent exposure effects from occurring.

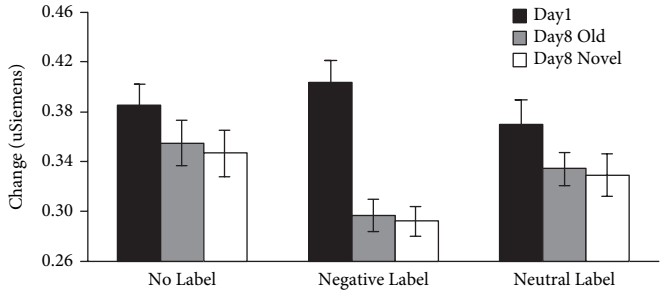

Fig. 13–4 Spider phobic skin conductance responses to spider images as a function of day and initial encoding condition. Higher bars indicate greater reactivity. For the labeling conditions (Negative Label, Neutral Label), the labels were present on Day 1, but on Day 8, pictures were presented without labels for all conditions.

A week after the first session, participants returned for a second session. On Day 8, participants were again shown the same pictures from Day 1 while SCR was measured; however, on Day 8, no words were shown for any of the conditions. By comparing SCR to pictures in each condition across the two sessions, we could determine the extent to which repeated exposure on Day 1 led to diminished SCR a week later and also whether the addition of affect labels enhanced this effect. As predicted, pictures that had been presented alone on Day 1 produced diminished SCRs on Day 8. This was also true for pictures that were presented with negative words on Day 1; however, pictures presented with neutral words on Day 1 only showed a trend in this direction. Critically, although both pictures shown alone and pictures shown with negative words showed diminished SCRs on Day 8, the reduction for the negative word condition was greater than the reduction for the no word condition.

This effect was replicated in a second study (Tabibnia et al., 2008a), examining the SCRs of individuals with spider fears to pictures of spiders. In this between-groups study, individuals saw pictures of spiders in one condition only (no words, negative words, or neutral words). In each condition, participants produced smaller SCRs to spider pictures on Day 8 than on Day 1 and replicating the first study, this effect was significantly greater in the negative words condition than the no words condition (*see* Fig. 13–4). Interestingly, the effects of the negative words shown on Day 1 generalized to new pictures of spiders that were not shown on Day 1 and had never been paired with words. Thus, these results suggest that pairing affect labels with repeated exposures of feared stimuli can lead to long-term reductions in the emotional responses to those stimuli.

More generally, these results point to the benefits of examining how specific symbolic processes unique to humans can benefit mental health processes. There has been a great deal of work in the past decade to translate the animal research on extinction processes into the human domain and demonstrating that these processes do translate from rodent to human. At the same time, humans have specific capacities that we do not share with other animals and these undoubtedly modulate the ways in which the lower processes operate within humans (Davey, 1992).

Social cognitive implications

Automaticity and Control

In addition to the applied clinical applications of disruption theory, this work also has important implications for both theory and methods within social cognition. First, the findings from this work suggest that our basic definitions of

automaticity and control, a core distinction within social cognition (Chaiken & Trope, 1999), need to be revisited (cf. Bargh, 1989). One of the gold standards for determining whether a process is automatic is to observe whether the process still occurs when the eliciting stimulus is presented subliminally (Monahan, Murphy, & Zajonc, 2000; Murphy & Zajonc, 1993). Thus, if a trait word is presented subliminally and influences subsequent personality judgments, all would agree that this represents automatic or implicit priming. A second standard that has been used has been the amount of time a mental process takes to occur. Generally speaking, the effects of a prime word on the processing of a second word that follows within 300 milliseconds of the prime word are thought to be automatic (Neely, 1977). Finally, processes that produce the same outputs when a person is under cognitive load (i.e., mental distraction usually caused by a concurrent task), as when there is no cognitive load, are also considered to be automatic (Gilbert, 1989).

By the first two of these definitions, the amygdala response to emotional images is an exemplary case of automaticity. Multiple studies have demonstrated that the amygdala responds to subliminal presentations of emotional images (Morris et al., 2000; Whalen et al., 1999) and also that the amygdala responds within 150 milliseconds of stimulus presentation. Clearly, no conscious mental resources are needed to produce the amygdala's response to emotional stimuli. Indeed, the race-matching task, which produced the greatest amount of amygdala activity in a comparison with race labeling (Lieberman et al., 2005), was performed at the same speed with a concurrent working memory task as without this task.

Nevertheless, when individuals are asked to process affect labels while looking at negative emotional images, the amygdala response either disappears or is significantly attenuated. Here, the presence of a particular kind of concurrent controlled processing task (i.e., affect labeling) modulates what would otherwise be an automatic response in the amygdala. This runs counter to the dogma of standard dual-process models that controlled processes cannot affect automatic processes. How could a process that can occur during subliminal presentations when an individual has no awarness at all of the eliciting stimulus possibly be prevented or attenuated by conscious processing?

Once a cognitive neuroscience approach to automatic and controlled social cognition is taken (Lieberman, 2007), the answer is actually quite straightforward. One possibility is that the amygdala performs its operations automatically as has often been supposed (Pasley et al., 2004). On this account, the amygdala in no way depends on cognitive resources or controlled processing to perform its computations. However, the amygdala receives inputs from various regions of prefrontal cortex (Ghashghei & Barbas, 2002), and the functional effect of some of these inputs may be inhibitory (Quirk et al., 2003; Rosenkranz & Grace, 2002). Affect labeling may interfere with amygdala processing not because they compete for a limited pool of cognitive resources (as is assumed to be the case for competing controlled processes) but because affect labeling just happens to activate a prefrontal region that has inhibitory inputs to the amygdala. Thus, processes internal to the amygdala may well be automatic, and yet at the same time, other brain structures may be capable of modulating or inhibiting these processes. On the one hand, this suggests that some individual neural mechanisms may follow the standard principles of automaticity, but on the other hand it suggests that at a system level, our understanding of automaticity and control may be far too simplistic.

Semantic versus Embodied Emotion

A second issue for social cognition is the use of word-and-picture primes in experimental studies. It is not uncommon for social psychological research to use word-and-picture primes interchangeably (e.g., Dasgupta & Greenwald, 2001; Devine, Plant, Amodio, Harmon-Jones, & Vance, 2002; Galinsky & Moskowitz, 2000; Lowery, Hardin, & Sinclair, 2001; Wittenbrink, Judd, & Park, 2001). This may be a result of assuming that there are unified representations in the mind and that any stimulus relevant to that mental construct is going to activate this

unified representation. From this perspective, it might seem that the same representation can be activated implicitly or explicitly, but ultimately the same representation is invoked. The cognitive neuroscience of memory has demonstrated that not only are there implicit and explicit memory processes (i.e., ways of using and invoking mental representations) but also that there are distinct neural mechanisms that retain different aspects of our past experience in qualitatively distinct representations (e.g., episodic, semantic, conditioning). In the context of affect labeling research, it seems that negative emotional stimuli can also activate distinct representations and processes depending to some extent on whether the eliciting stimuli are words or pictures. Negatively valenced pictures reliably activate the amygdala and also lead to SCR increases, suggesting that *embodied* emotional processes are invoked. Alternatively, negatively valenced words produce neither of these effects and instead activate RVLPFC. Thus, it is possible that these words are producing thoughts about emotion but are not producing, or may even be inhibiting, basic emotional responses. In a pure social cognition task with word-only primes, this effect may be overlooked as negatively valenced words will presumably activate a *semantic* network of emotion representations (Robinson & Clore, 2002). It appears that it would be a mistake, however, to infer from the activation of this semantic network that more basic and embodied emotional processes have also been activated. Although this distinction has yet to be fully fleshed out, it does suggest that we may not be priming what we think we're priming in affect priming studies.

Conclusion

Numerous philosophers and psychologists have noted over the years that thinking about affect has the capacity to alter and even dampen the affect that is being thought about. This has been used to good effect in various forms of therapy, from formal psychotherapies to informal social support networks in which people talk about their feelings with friends. The reason that putting feelings into words helps has remained elusive and somewhat mystical. The work presented here describes a neurocognitive process focused on RVLPFC that provides the beginnings of an answer. Putting feelings into words activates a region of the brain that is capable of inhibiting various aspects of immediate experience, including affective distress. Although we cannot say why the brain evolved such that putting feelings into words has this neurocognitive effect, knowing that it does allows us to probe various aspects of this process in the future and examine its contribution to various social and affective experience in healthy and clinical populations.

Acknowledgments

Although the references have been updated, this chapter was written a few years ago and with fast developments in social cognitive neuroscience, more recent reviews can be found in Berkman and Lieberman (2009), Cohen and Lieberman (2010), and Lieberman (2010).

References

Abe, N, Suzuki, M., Tsukiura, T., et al. (2006). Dissociable roles of prefrontal and anterior cingulate cortices in deception. *Cerebral Cortex, 16*, 192–199.

Analayo (2003). *Satipatthana: The Direct Path to Realization*. Birmingham, UK: Windhorse publications.

Anderson, I. M., Clark, L., Elliott, R., Kulkarni, B., Williams, S. R., & Deakin, J. F. W. (2002). 5-HT2c receptor activation by m-chlorophenylpiperazine detected in humans with fMRI. *Neuroreport, 13*, 1547–1551.

Aron, A. R., Fletcher, P. C., Bullmore, E. T., Sahakian, B. J., & Robbins, T. W. (2003). Stop-signal inhibition disrupted by damage to right inferior frontal gyrus in humans. *Nature Neuroscience, 6*, 115–116.

Asahi, S., Okamoto, Y., Okada, G., Yamawaki, S., & Yokota, N. (2004). Negative correlation between right prefrontal activity during response inhibition and impulsiveness: A fMRI study. *European Archives of Clinical Neuroscience, 254*, 245–251.

Augustine, J. R. (1996). Circuitry and functional aspects of the insular lobe in primates

including humans. *Brain Research Reviews, 22*, 229–244.

Baikie, K. A. & Wilhelm, K. (2005). Emotional and physical health benefits of expressive writing. *Advances in Psychiatric Treatment, 11*, 338–346.

Bargh, J. A. (1989). Conditional automaticity: Varieties of automatic influence in social perception and cognition. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended Thought* (pp. 3–51). New York: Guilford.

Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology, 71*, 230–244.

Beauregard, M., & Levesque, J. (2006). Functional magnetic resonance imaging investigation of the effects of neurofeedback training on the neural bases of selective attention and response inhibition in children with attention-deficit/hyperactivity disorder. *Applied Psychophysiological Biofeedback, 31*, 3–20.

Beauregard, M., Levesque, J., & Bourgouin, P. (2001). Neural correlates of conscious self-regulation of emotion. *Journal of Neuroscience, 21*, 6993–7000.

Bennedetti, F., Mayberg, H. S., Wager, T. D., Stohler, C. S., & Zubieta, J. K. (2005). Neurobiological mechanisms of the placebo effect. *Journal of Neuroscience, 25*, 10,390–10,402.

Berkman, E., & Lieberman, M. D. (2009). Using neuroscience to broaden emotion regulation: Theoretical and methodological considerations. *Social and Personality Psychology Compass, 3*, 475–493.

Blasi, G., Goldberg, T. E., Weickert, T., et al. (2006). Brain regions underlying response inhibition and interference monitoring and suppression. *European Journal of Neuroscience, 23*, 1658–1664.

Booth, R. J., Petrie, K. J., & Pennebaker, J. W. (1997). Changes in circulating lymphocyte numbers following emotional disclosure: evidence of buffering? *Stress Medicine, 13*, 23–29.

Britton, J. C., Taylor, S. F., Sudheimer, K. D., & Liberzon, I. (2006). Facial expressions and complex IAPS pictures: Common and differential networks. *Neuroimage, 31*, 906–919.

Broderick, J. E., Junghaenel, D. U., & Schwartz, J. E. (2005). Written emotional expression produces health benefits in fibromyalgia patients. *Psychosomatic Medicine, 67*, 326–334.

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

Chaiken, S., & Trope, Y. (1999). *Dual-process Theories in Social Psychology*. New York: Guilford Press.

Chambers, C. D., Bellgrove, M. A., Stokes, M. G., et al. (2006). Executive "brake failure" following deactivation of human frontal lobe. *Journal of Cognitive Neuroscience, 18*, 444–455.

Cohen, J. R. & Lieberman, M. D. (2010). The common neural basis of exerting self-control in multiple domains. To appear in Y. Trope, R. Hassin, & K. N. Ochsner (eds.), *Self-control* (pp. 141–160). Oxford University Press.

Critchley, H., Daly, E., Phillips, M., et al. (2000). Explicit and implicit neural mechanisms for processing of social information from facial expressions: A functional magnetic resonance imaging study. *Human Brain Mapping, 9*, 93–105.

Crockett, M. J., Lieberman, M. D., & Tabibnia, G. (unpublished manuscript). Affect labeling attenuates skin conductance responses to emotionally evocative faces in neurotics. University of California, Los Angeles.

Cunningham, W. A., Espinet, S. D., Deyoung, C. G., & Zelazo, P. D. (2006). Attitudes to the right- and left: Frontal ERT asymmetries associated with stimulus valence and processing goals. *NeuroImage, 28*, 827–834.

Cunningham, W. A., Johnson, M. K., Gatenby, J. C., Gore, J. C., & Banaji, M. R. (2003). Neural components of social evaluation. *Journal of Personality and Social Psychology, 85*, 639–649.

Cunningham, W. A., Johnson, M. K., Raye, C. L., Gatenby, J. C., Gore, J. C., & Banaji, M. R. (2004). Separable neural components in the processing of black and white faces. *Psychological Science, 15*, 806–813.

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

Dandoy, A. C., & Goldstein, A. G. (1990). The use of cognitive appraisal to reduce stress reactions: A replication. *Journal of Social Behavior and Personality, 5*, 275–285.

Dasgupta, N., & Greenwald, A. G. (2001). On the malleability of automatic attitudes: Combating automatic prejudice with images of admired

and disliked individuals. *Journal of Personality and Social Psychology, 81*, 800–814.

Davey G. C. L. (1992): Classical conditioning and the acquisition of human fears and phobias: A review and synthesis of the literature. *Advances in Behavioral Research and Therapy, 14*, 29–66.

Del-Ben, C. M., Deakin, J. F. W., Mckie, S., et al. (2005). The effect of citalopram pretreatment on neuroanl responses to neuropsychological tasks in normal volunteers: An fMRI study. *Neuropsychopharmacology, 30*, 1724–1734.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56*, 5–18.

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.

Dolcos, F., LaBar, K. S., & Cabeza, R. (2004). Dissociable effects of arousal and valence on prefrontal activity indexing emotional evaluation and subsequent memory: An event-related fMRI study. *Neuroimage, 23*, 64–74.

Durston, S., Mulder, M., Casey, B. J., Ziermans, T., & van Engeland, H. (2006). Activation in ventral prefrontal cortex is sensitive to genetic vulnerability for attention-deficit hyperactivity disorder. *Biological Psychiatry, 60*, 1062–1070.

Eisenberger, N. I. (2006). Identifying the neural correlates underlying social pain: Implications for developmental processes. *Human Development, 49*, 273–293.

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.

Eisenberger, N. I., Way, B., Taylor, S. E., Welch, W. T., Lieberman, M. D. (2007). Understanding genetic risk for aggression: Clues from the brain's response to social exclusion. *Biological Psychiatry, 61*, 1100–1108.

Fan, J., Flombaum, J. I., McCandiss, B. D., Thomas, K. M., & Posner, M. I. (2003). Cognitive and brain consequences of conflict. *NeuroImage, 18*, 42–57.

Foa, E. B., & Kozak, M. J. (1986). Emotional processing of fear: Exposure to corrective information. *Psychology Bulletin, 99*, 20–35.

Francis, M. E., & Pennebaker, J. W. (1992). Putting stress into words. The impact of writing on physiological, absentee, and self-reported emotional well-being measures. *American Journal of Health Promotion, 6*, 280–287.

Galinsky, A. D., & Moskowitz, G. B. (2000). Perspective-taking: Decreasing stereotype expression, stereotype accessibility, and in-group favortism. *Journal of Personality and Social Psychology, 78*, 708–724.

Garavan, H., Ross, T. J., & Stein, E. A. (1999). Right hemispheric dominance of inhibitory control: An event-related functional MRI study. *Proceedings of the National Academy of Science, 96*, 8301–8306.

Ghashghaei, H. T., & Barbas, H. (2002). Pathways for emotion: Interactions of prefrontal and anterior temporal pathways in the amygdala of the rhesus monkey. *Neuroscience, 115*, 1261–1279.

Gilbert, D. T. (1989). Thinking lightly about others. Automatic components of the social inference process. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended Thought* (pp. 189–211). New York: Guilford.

Goel, V., & Dolan, R. J. (2003). Reciprocal neural response within lateral and ventral medial prefrontal cortex during hot and cold reasoning. *Neuroimage, 20*, 4314–4321.

Gorno-Tempini, M. L., Pradelli, S., Serafini, M., et al. (2001). Explicit and incidental facial expression processing. An fMRI study. *NeuroImage, 14*, 465–473.

Greenberg, M. A., Wortman, C. B., & Stone, A. A. (1996). Emotional expression and physical health: revising traumatic memories or fostering self-regulation? *Journal of Personality and Social Psychology, 71*, 588–602.

Gross, J. J. (1998). Antecedent- and response-focused emotion regulation: Divergent consequences for experience, expression, and physiology. *Journal of Personality and Social Psychology, 74*, 227–237.

Gur, R. C., Schroeder, L., Turner, T., et al. (2002). Brain activation during facial emotion processing. *NeuroImage, 16*, 651–662.

Hariri, A. R., Bookheimer, S. Y., & Mazziotta, J. C. (2000). Modulating emotional response: Effects of a neocortical network on the limbic system. *NeuroReport, 11*, 43–48.

Hazeltine, E., Bunge, S. A., Scanlon, M. D., & Gabrieli, J. D. E. (2003). Material-dependent and material-indpendent selection processes in

the frontal and parietal lobes: An event-related fMRI investigation of response competition. *Neuropsychologia*, 41, 1208–1217.

Hazeltine, E., Poldrack, R., & Gabrieli, J. D. E. (2000). Neural activation during response competition. *Journal of Cognitive Neuroscience, 12*, 118–129.

Hemenover, S. H. (2003). The good, the bad, and the healthy: Impacts of emotional disclosure of trauma on resilient self-concept and psychological distress. *Personality and Social Psychology Bulletin, 29*, 1236–1244.

Horn, N. R., Dolan, M. Elliott, R., Deakin, J. F. W., & Woodruff, P. W. R. (2003). Response inhibition and impulsivity: An fMRI study. *Neuropsychologia, 41*, 1959–1966.

Hornak, J., Rolls, E. T., & Wade, D. (1996). Face and voice expression identification in patients with emotional and behavioural changes following ventral frontal lobe damage. *Neuropsychologia, 34*, 247–261.

Jahn, K., Deutschlander, A., Stephan, T., Strupp, M., Wiesmann, M., & Brandt, T. (2004). Brain activation patterns during imagined stance and locomotion in functional magnetic resonance imaging. *Neuroimage, 22*, 1722–1731.

James, W. (1890/1950). *The Principles of Psychology*. New York: Dover.

Kalisch, R., Wiech, K., Critchley, H. D., et al. (2005). Anxiety reduction through detachment: Subjective, physiological, and neural effects. *Journal of Cognitive Neuroscience, 17*, 874–883.

Kawashima, R., Satoh, K., Itoh, H., et al. (1996). Functional anatomy of GO/NO-GO discrimination and response selection—a PET study in man. *Brain Research, 728*, 79–89.

Kemmotsu, N., Villalobos, M. E., Gaffrey, M. S., Courchesne, E., & Muller, R. A. (2005). Activity and functional connectivity of inferior frontal cortex associated with response conflict. *Cognitive Brain Research, 24*, 335–342.

Konishi, S., Nakajima, K., Uchida, I., Kikyo, H., Kameyama, M., & Miyashita, Y. (1999). Common inhibitory mechanism in human inferior prefrontal cortex revealed by event-related functional MRI. *Brain, 122*, 981–999.

Kuchinke, L., Jacobs, A. M., Grubich, C., Vo, M. L., Conrad, M., & Hermann, M. (2005). Incidental effects of emotional valence in single word processing. An fMRI study. *NeuroImage, 28*, 1022–1032.

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.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1999). *International Affective Picture System (IAPS): Instruction Manual and Affective Ratings*. Gainsville: University of Florida, The Center for Research in Psychophysiology.

Lange, K., Williams, L. M., Young, A. W., et al. (2003). Task instructions modulate neural response to fearful facial expressions. *Biological Psychiatry, 53*, 226–232.

Lazarus, R. S., & Alfert, E. (1964). Short-circuiting of threat by experimentally altering cognitive appraisal. *Journal of Abnormal and Social Psychology, 69*, 195–205.

Lazarus, R. S., Opton, E. M., Nomikos, M. S., & Rankin, N. O. (1965). The principle of short-circuiting of threat: Further evidence. *Journal of Personality, 33*, 622–635.

LeDoux, J. E. (1996). *The Emotional Brain: The Mysterious Underpinnings of Emotional Life*. New York: Simon & Schuster.

Lepore, S. J., & Smyth, J. M. (2002). *The Writing Cure: How Expressive Writing Promotes Health and Emotional Well-being*. Washington D.C.: American Psychological Association.

Leung, H., Skudlarski, P., Gatenby, J. C., Peterson, B. S., & Gore, J. C. (2000). An event-related functional MRI study of the Stroop color word interference task. *Cerebral Cortex, 10*, 552–560.

Levesque, J., Eugene, F., Joanette, Y., et al. (2003). Neural circuitry underlying voluntary suppression of sadness. *Biological Psychiatry, 53*, 502–510.

Liddle, P. F., Kiehl, K. A., & Smith, A. M. (2001). Event-related fMRI study of response inhibition. *Human Brain Mapping, 12*, 100–109.

Lieberman, M. D. (2007). The X- and C-sytems: The neural basis of reflexive and reflective social cognition. In E. Harmon-Jones & P. Winkelman (eds.), *Fundamentals of Social Neuroscience* (pp. 290–315). New York: Guilford.

Lieberman, M. D. (2010). Social cognitive neuroscience. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *Handbook of Social Psychology* (5th ed., pp. 143–193). New York, NY: McGraw-Hill.

Lieberman, M. D., Eisenberber, N. I., & Crockett, M. J. (unpublished manuscript). An fMRI study of automatic behavior: Comparing

ideomotor and disruption accounts. University of California, Los Angeles.

Lieberman, M. D., Eisengerger, N. I., Crockett, M. J., Tom, S. M., Pfeifer, J. H., & Way, B. M. (2007). Putting feelings into words: Affect labeling disrupts amygdala activity to affective stimuli. *Psychological Science, 18*, 421–428.

Lieberman, M. D., Hariri, A., Jarcho, J. J., Eisenberger, N. I., & Bookheimer, S. Y. (2005). An fMRI investigation of race-related amygdala activity in African-American and Caucasian-American individuals. *Nature Neuroscience, 8*, 720–722.

Lieberman, M. D., Jarcho, J. M., Berman, S., et al. (2004). The neural correlates of placebo effects: A disruption account. *NeuroImage, 22*, 447–455.

Livingston, R. W., & Brewer, M. B. (2002). What are we really priming? Cue-based versus category-based processing of facial stimuli. *Journal of Personality and Social Psychology, 82*, 5–18.

Lowery, B. S., Hardin, C. D., & Sinclair, S. (2001). Social influence on automatic racial prejudice. *Journal of Personality and Social Psychology, 81*, 842–855.

Luan Phan, K., Magalhaes, A., Ziemlewicz, T. J., Fitzgerald, D. A., Green, C., & Smith, W. (2005). Neural correlates of telling lies: A functional magnetic resonance imaging study at 4 tesla. *Academic Radiology, 12*, 164–172.

Lyubomirsky, S., Sousa, L., & Dickerhoof, R. (2006). The costs and benefits of writing, talking, and thinking about life's triumphs and defeats. *Journal of Personality and Social Psychology, 90*, 692–708.

Martin, J. H. (1996). *Neuroanatomy* (2nd ed.). Stamford, CT: Appleton & Lange.

Matthews, S. C., Simmons, A. N., Arce, E., & Paulus, M. P. (2005). Dissociation of inhibition from error processing using a parametric inhibitory task during functional magnetic resonance imaging. *Neuroreport, 16*, 755–760.

McDonald, A. J., Mascagni, F., & Guo, L. (1996). Projections of the medial and lateral prefrontal cortices to the amygdala: A *phaseolus vulgaris* leucoagglutinin study in the rat. *Neuroscience, 71*, 55–75.

McGuire, K. M. B., Greenberg, M. A., & Gevirtz, R. (2005). Autonomic effects of expressive writing in individuals with elevated blood pressure. *Journal of Health Psychology, 10*, 197–209.

Monahan, J. L., Murphy, S. T., & Zajonc, R. B. (2000). Subliminal mere exposure: Specific, general, and diffuse effects. *Psychological Science, 11*, 462–466.

Morris, J. S., Ohman, A., & Dolan, R. J. (1999). A subcortical pathway to the right amygdala mediating "unseen" fear. *Proceedings of National Academy of Science, USA, 96*, 1680–1685.

Morris, J. S., DeGelder, B., Weiskrantz, L., & Dolan, R. J. (2001). Differential extrageniculostriate and amygdala responses to presentation of emotional faces in a cortically blind field. *Brain, 124*, 1241–1252.

Murphy, S. T., & Zajonc, R. B. (1993) Affect, cognition, and awareness: Affective priming with optimal and suboptimal stimulus exposures. *Journal of Personality and Social Psychology, 64*, 723–739.

Nakamura, K., Kawashima, R., Ito, K., et al. (1999). Activation of the right inferior frontal cortex during assessment of facial emotion. *Journal of Neurophysiology, 82*, 1610–1614.

Naliboff, B. D., Berman, S., Suyenobu, B., et al. (2006). Longitudinal change in perceptual and brain activation response to visceral stimuli in irritable bowel syndrome patients. *Gastroenterology, 131*, 352–365.

Narumoto, J., Yamada, H., Iidaka, T., et al. (2000). Brain regions involved in verbal or non-verbal aspects of facial emotion recognition. *NeuroReport, 11*, 2571–2576.

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited capacity attention. *Journal of Experimental Psychology: General, 106*, 226–254.

Nolen-Hoeksema, S. (2000). The role of rumination in depressive disorders and mixed anxiety/depressive symptoms. *Journal of Abnormal Psychology, 109*, 504–511.

Nomura, M., Iidaka, T., Kakehi, K., et al. (2003). Frontal lobe networks for effective processing of ambiguously expressed emotions in humans. *Neuroscience Letters, 348*, 113–116.

Nosek, B.A., Banaji, M. R., & Greenwald, A. G. (2002). Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dynamics: Theory, Research, and Practice, 6*, 101–115.

Nunez, J. M., Casey, B. J., Egner, T., Hare, T., & Hirsch, J. (2005). Intentional false responding shares neural substrates with response conflict and cognitive control. *Neuroimage, 25*, 267–277.

Ochsner, K. N., Bunge, S. A., Gross, J. J., & Gabrieli, J. D. E. (2002). Rethinking feelings: An fMRI study of the cognitive regulation of emotion. *Journal of Cognitive Neuroscience, 14*, 1215–1229.

Ochsner, K. N., Ray, R. D., Cooper, J. C., et al. (2004). For better or for worse: Neural systems supporting the cognitive down- and up-regulation of negative emotion. *Neuroimage, 23*, 483–499.

Park, C. L., & Blumberg, C. J. (2002). Disclosing trauma through writing: Testing the meaning-making hypothesis. *Cognitive Therapy and Research, 26*, 597–616.

Pasley, B. N., Mayes, L. C., & Schultz, R. T. (2004). Subcortical discrimination of unperceived objects during binocular rivalry. *Neuron, 42*, 163–172.

Pennebaker, J. W. (1997). Writing about emotional experiences as a therapeutic process. *Psychological Science, 8*, 162–166.

Pennebaker, J. W., & Beall, S. K. (1986). Confronting a traumatic event. Toward an understanding of inhibition and disease. *Journal of Abnormal Psychology, 95*, 274–281.

Petrovic, P., Dietrich, T., Fransson, P., Andersson, J., Carlsson, & Ingvar, M. (2005). Placebo in emotional processing—induced expectations of anxiety relief activate a generalized modulatory network. *Neuron, 46*, 957–969.

Petrovic, P., Kalso, E., Petersson, K. M., & Ingvar, M. (2002). Placebo and opioid analgesia imaging a shared neuronal network. *Science, 295*, 1737–1740.

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

Quirk, G. J., Likhtik, E., Pelletier, J. G., & Pare, D. (2003). Stimulation of medial prefrontal cortex decreases the responsiveness of central amygdala output neurons. *Journal of Neuroscience, 23*, 8800–8807.

Richards, J. M., & Gross, J. J. (2000). Emotion regulation and memory: The cognitive costs of keeping one's cool. *Journal of Personality and Social Psychology, 79*, 410–424.

Robinson, M. D., & Clore, G. L. (2002). Belief and feeling: evidence for an accessibility model of emotional self-report. *Psychological Bulletin, 128*, 934–960.

Robinson, M. D., & Wilkowski, B. M. (2006). Loving, hating, vacillating: Agreeableness, implicit self-esteem, and neurotic conflict. *Journal of Personality, 74*, 935–977.

Rosenkranz, J. A., & Grace, A. A. (2002). Cellular mechanisms of infralimbic and prelimbic prefrontal cortical inhibition and dopaminergic modulation of basolateral amygdala neuronsin vivo. *Journal of Neuroscience, 22*, 324–337.

Royet, J. P., Plailly, J., Delon-Martin, C., Kareken, D. A., & Segebarth, C. (2003). fMRI of emotional responses to odors: Influence of hedonic valence and judgments, handedness, and gender. *NeuroImage, 20*, 713–728.

Rubia, K., Lee, F., Cleare, A. J., et al. (2005). Tryptophan depletion reduces right inferior prefrontal activation during response inhibition in fast, event-related fMRI. *Psychopharmacology, 179*, 791–803.

Rubia, K., Overmeyer, S., Taylor, E., et al. (1999). Hypofrontality in attention deficit hyperactivity disorder during higher-order motor control: A study with functional MRI. *American Journal of Psychiatry, 156*, 891–896.

Rubia, K., Smith, A. B., Brammer, M. J., & Taylor, E. (2003). Right inferior prefrontal cortex mediates response inhibition while mesial prefrontal cortex is responsible for error detection. *NeuroImage, 20*, 351–358.

Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2003). Theneural basis of economic decision-making in the Ultimatum Game. *Science, 300*, 1755–1758.

Schaefer, A., Collette, F., Philippot, P., et al. (2003). Neural correlates of "hot" and "cold" emotional processing: A multilevel approach to the functional anatomy of emotion. *NeuroImage, 18*, 938–949.

Shaw, P., Bramham, J., Lawrence, E. J., Morris, R., Baron-Cohen, S., & David, A. S. (2005). Differential effects of lesions of amygdala and prefrontal cortex on recognizing facial expressions of complex emotions. *Journal of Cognitive Neuroscience, 17*, 1410–1419.

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

Smyth, J. M., Stone, A. A., & Hurewitz, A. (1999). Effects of writing about stressful experiences on symptom reduction in patients with asthma or rheumatoid arthritis. A randomized trial. *Journal of the American Medical Association, 281*, 1304–1309.

Speisman, J. C., Lazarus, R. S., Mordkoff, A. M., & Davison, L. A. (1964). Experimental reduction of stress based on ego-defense theory. *Journal of Abnormal and Social Psychology, 68*, 367–380.

Spence, S. A., Farrow, T. F. D., Herford, A. E., Wilkinson, I. D., Zheng, Y., & Woodruff, P. W. R. (2001). Behavioural and functional anatomical correlates of deception in humans. *Neuroreport, 13*, 2849–2852.

Spinoza, B. (1675/2000). *Ethics*. New York: Oxford University Press.

Stanton, A. L., Danoff-Burg, S., Sworowski, L. A., et al. (2002). Randomized, controlled trial of written emotional expression and benefit finding in breast cancer patients. *Journal of Clinical Oncology, 15*, 4160–4168.

Tabibnia, G., Lieberman, M. D., & Craske, M. (2008a). The lasting effect of words on feelings: Words may facilitate exposure effects to threatening images. *Emotion, 8*, 307–317.

Tabibnia, G., Satpute, A. B., & Lieberman, M. D. (2008b). The sunny side of fairness: Preference for fairness activates reward circuitry (and disregarding unfairness activates self-control circuitry). *Psychological Science, 19*, 339–347.

Taylor, S. F., Phan, K. L., Decker, L. R., & Liberzon, I. (2003). Subjective rating of emotionally salient stimuli modulates neural activity. *NeuroImage, 18*, 650–659.

Vogt, B. A. & Pandya, D. N. (1987). Cingulate cortex of the rhesus monkey: II. Cortical afferents. *Journal of Comparative Neurology, 262*, 271–289.

Vollm, B., Richardson, P., McKie, S., Elliott, R., Deakin, J. F. W., & Anderson, I. M. (2006). Serotonergic modulation of neuronal responses to behavioural inhibition and reinforcing stimuli: An fMRI study in healthy volunteers. *European Journal of Neuroscience, 23*, 552–560.

Vuilleumier, P., Armony, J. L., Driver, J., & Dolan, R. J. (2001). Effects of attention and emotion on face processing the human brain: An event-related fMRI study. *Neuron, 30*, 829–841.

Wager, T. D., Rilling, J. K., Smith, E. E., et al. (2004). Placebo-induced changes in fMRI in the anticipation and experience of pain. *Science, 303*, 1162–1167.

Wegner, D. M., Erber, R., & Zanakos, S. (1993). Ironic processes in the mental control of mood and mood-related thought. *Journal of Personality and Social Psychology, 65*, 1093–1104.

Wegner, D. M., Shortt, J. W., Blake, A. W., & Page, M. S. (1990). The suppression of exciting thoughts. *Journal of Personality and Social Psychology, 58*, 409–418.

Whalen, P. J., Rauch, S. L., Etcoff, N. L., McInerney, S. C., Lee, M. B., & Jenike, M. A. (1998). Masked presentations of emotional facial expressions modulate amygdala activity without explicit knowledge. *Journal of Neuroscience, 18*, 411–418.

Wildgruber, D., Hertrich, I., Riccker, A., et al. (2004). Distinct frontal regions subserve evaluation of linguistic and emotional aspects of speech intonation. *Cerebral Cortex, 14*, 1384–1389.

Wildgruber, D., Riecker, A., Hertrich, I., et al. (2005). Identification of emotional intonation evaluated by fMRI. *NeuroImage, 24*, 1233–1241.

Williams, K. D. (in press). Ostracism. *Annual Review of Psychology, 58*.

Wilson, T. D., & Schooler, J. W. (1991). Thinking too much: Introspection can reduce the quality of preferences and decisions. *Journal of Personality and Social Psychology, 60*, 181–192.

Wilson, T. D., Lisle, D. J., Schooler, J. W., Hodges, S. D., Klaaren, K. J., & LaFleur, S. J. (1993). Introspecting about reasons can reduce post-choice satisfaction. *Personality and Social Psychology Bulletin, 19*, 331–339.

Wittenbrink, B., Judd, C. M., & Park, B. (2001). Spontaneous prejudice in context: Variability in automatically activated attitudes. *Journal of Personality and Social Psychology, 81*, 815–827.

AQ1

AQ1: Please update the reference Williams.