NEURAL AND BEHAVIORAL EFFECTS OF REGULATING EMOTIONAL RESPONSES TO ERRORS DURING AN IMPLICIT RACIAL BIAS TASK

A Thesis

presented to

the Faculty of the Graduate School

at the University of Missouri-Columbia

In Partial Fulfillment

of the Requirements for the Degree

Master of Arts

by

MEREDITH P. JOHNSON

Dr. Bruce D. Bartholow, Thesis Supervisor

MAY 2015
The undersigned, appointed by the dean of the Graduate School, have examined the thesis entitled

NEURAL AND BEHAVIORAL EFFECTS OF REGULATING EMOTIONAL RESPONSES TO ERRORS DURING AN IMPLICIT BIAS TASK

presented by Meredith Phinizy Johnson,

a candidate for the degree of master of psychology,

and hereby certify that, in their opinion, it is worthy of acceptance.

______________________________
Professor Bruce Bartholow

______________________________
Professor John Kerns

______________________________
Professor Elizabeth Behm-Morawitz
DEDICATION

I dedicate this to my father, Dr. Stewart Phinizy Johnson, in whose doctoral footsteps I am following. He is my academic beacon of light. Also, I dedicate this to my mother, Jane Kestenbaum, who has demonstrated that hard work pays off. Lastly, I would like to dedicate this to my younger sister, Emma Louise Johnson, because I think she is great.
ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Dr. Bruce Bartholow, for his guidance and support throughout all stages of the development and execution of the work presented here. I would also like to thank Drs. John Kerns and Lissa Behm-Morawitz for serving on my thesis committee, and for providing me with thoughtful comments and helpful feedback.

I would further like to thank the undergraduate research assistants who helped collect the data. I am also indebted to the undergraduate research participants, without whom, this project may have never come to fruition.

I owe a very special thanks to the members of the Bartholow Lab for enduring countless presentations of the data, and for statistical guidance, intellectual input, and emotional support along the way.

Finally, I would like to thank Gatsby for staying up late with me and editing drafts of this manuscript during the final stages of the revision process. Also, I owe a huge thanks to Tyler Levsen for being my first pilot subject, enthusiastically listening to practice presentations of this data, providing me with sustenance when my academics took priority, and for being a bottomless well of emotional support.
TABLE OF CONTENTS

ACKNOWLEDGEMENTS ........................................................................................................ ii
LIST OF TABLES .................................................................................................................. iv
LIST OF FIGURES ................................................................................................................. v
ABSTRACT .......................................................................................................................... vi

Chapter

1. INTRODUCTION ............................................................................................................. 1

2. STUDY 1 .......................................................................................................................... 10
   a. METHODS .............................................................................................................. 10
   b. RESULTS .............................................................................................................. 16
   c. DISCUSSION ......................................................................................................... 24

3. STUDY 2 .......................................................................................................................... 29
   a. METHODS .............................................................................................................. 30
   b. RESULTS .............................................................................................................. 32
   c. DISCUSSION ......................................................................................................... 36

4. GENERAL DISCUSSION .............................................................................................. 41

TABLES ................................................................................................................................. 50
FIGURES ............................................................................................................................... 52
REFERENCES ....................................................................................................................... 56

APPENDIX

1. SUPPLEMENTARY RESULTS, DISCUSSION, AND REFERENCES .......................... 64
   Block order effect
   Error Positivity

2. SUPPLEMENTARY FIGURES ......................................................................................... 80
LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Within-Subjects (Study 1): Mean Accuracy Rates and Response Times in the Weapons Identification Task (WIT)</td>
<td>50</td>
</tr>
<tr>
<td>2. Between-Subjects (Studies 1 and 2): Mean Accuracy Rates and Response Times in the WIT</td>
<td>51</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Study 1 performance across block instructions (within subjects)</td>
<td>52</td>
</tr>
<tr>
<td>2. Within- and between-subjects comparison of the effect of emotion regulation strategies on the ERN</td>
<td>53</td>
</tr>
<tr>
<td>3. Between-subjects comparison of block instructions on accuracy rates and automatic and controlled processes</td>
<td>55</td>
</tr>
</tbody>
</table>

**Supplementary Figure**

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Comparison of $P_E$ across blocks during Study 1 (within-subjects) and the first block during Studies 1 and 2 (between-subjects)</td>
<td>80</td>
</tr>
<tr>
<td>2. The effect of block order on suppress and reappraise</td>
<td>81</td>
</tr>
<tr>
<td>3. The effect of block order on the ERN</td>
<td>83</td>
</tr>
</tbody>
</table>
NEURAL AND BEHAVIORAL EFFECTS OF REGULATING EMOTIONAL RESPONSES TO ERRORS DURING AN IMPLICIT RACIAL BIAS TASK

Meredith P. Johnson

Dr. Bruce Bartholow, Thesis Supervisor

ABSTRACT

Affect regulation plays a key role in several theories of prejudice reduction. Here, we tested whether engaging in emotion regulation strategies (ERSs) while performing an implicit racial bias task (Weapons Identification Task; WIT) would alter neural and behavioral manifestations of bias. Participants either suppressed or reappraised in a positive light the distress associated with making errors during the WIT while EEG was recorded. Originally, we hypothesized that if participants experienced less distress when they made errors, they would be less motivated to correct their behavior. We predicted this would reduce accuracy and increase expression of racial bias. Furthermore, we expected that the error-related negativity (ERN), hypothesized to originate from the anterior cingulate cortex (ACC) and reflect distress over errors, would be reduced during ERSs. However, contrary to initial predictions, results indicated that engaging in either ERS actually did not impact performance. Additionally, suppress altered the typical neural pattern observed during the WIT, such that the ERN following errors indicative of racial bias was no different from other types of errors, suggesting that suppress reduced signs of bias-related compunction. On the other hand, reappraisal was found to increase ERNs following errors indicative of racial bias. This may reflect the recruitment of overlapping cingulate cortical regions during reappraisal that are typically utilized during error monitoring while performing the WIT.
Introduction

Although racial inequality has been declining in recent decades (Fischer & Hout, 2006), the disadvantages of being a racial minority in the US still pose a major threat to this democracy’s status as a land of equal opportunity and equal rights for all citizens. It is believed that racial prejudice – be it conscious thoughts, or subtle feelings beyond the scope of awareness – drives individual behavior and decision-making in such a way that gives rise to the present state of racial inequality (e.g., Payne & Cameron, 2010).

Preferential evaluations and treatment towards others who share one’s group membership is what Tajfel and Turner (1986) described as in-group bias. They make the distinction between implicit and explicit intergroup conflicts, noting that implicit conflicts exist without official institutionalization or endorsement of the conflict. Evidence of such conflicts is typically found in group differentiations, for which there appears to be “no reason” (e.g., racial disparity in income; Tajfel & Turner, 1986). For instance, one sociological quasi-experiment found that White males were twice as likely to receive entry-level job offers as equally qualified Black males (Pager, Western, & Bonikowski, 2009). This example of implicit racial intergroup conflict illustrates how unintentional, unconscious racial bias contributes to racial inequality.

The unconscious type of bias discussed above is commonly referred to as implicit bias. It is defined as a preference that arises automatically during earlier stages of perception before more reflective (conscious), higher levels of processing have had time to take place (Cunningham, Raye, & Johnson, 2004). Previous research has found that implicit attitudes can be more predictive of actual behavior (Greenwald & Banaji, 1995). Numerous studies have demonstrated that White Americans tend to have more negative automatic evaluations of Black Americans and other Non-White ethnic groups relative to
other members of their own racial group (Dasgupta, McGhee, Greenwald, & Banaji, 2000; Phelps et al., 2000; Nosek et al., 2007). Implicit bias is influenced and formulated by past experiences that may be introspectively inaccessible at the time that the relevant judgment, evaluation or decision is made (Greenwald & Banaji, 1995). In contrast, explicit bias is a personal preference that an individual is consciously aware of having. Especially in regards to controversial or sensitive topics, these preferences may be influenced by social norms and cultural ideology. They are thought to represent a more controlled, reflective level of processing.

It is well established that implicit and explicit attitudes are dissociable from one another. Explicit attitudes do not always reflect or correspond to implicit attitudes (Nosek, 2007). One study found that implicit racial attitudes (based on a task measuring implicit preference for Whites or Blacks) were only mildly, positively related to explicit racial attitudes ($r=0.27$; Nosek, et al., 2007). The discordance between implicit and explicit racial attitudes has important implications for society as a whole. It has been argued that the persistence of racial inequality in the US (during a time when explicit forms of racial bias are now heavily stigmatized) is driven in part by behavioral manifestations of implicit racial bias (e.g. Payne & Cameron, 2010). There exists a plethora of research examining these types of behaviors, which have been described as very subtle and automatic, usually occurring outside of conscious awareness. Similar to implicit attitudes, these subtleties in behavior are unintentional.

In addition to subtle behaviors, implicit biases are also exhibited in the form of affective states. Contact with members of out-groups has been associated with negative affective responses, like anxiety (Stephan, Stephan, & Gudykunst, 1999). This is
reflected in physiological and neurological anxiety and fear responses. For example, Rankin and Campbell (1955) demonstrated that, despite their self-reported positive attitudes toward Blacks, brief physical contact with a Black experimenter increased the skin conductance response of White participants. The galvanic skin response (GSR) is a signal of physiological arousal and has been strongly correlated with anxiety. Brain imaging studies have shown that activation in the region known as the amygdala is greater for White participants while viewing Black faces compared to White faces (e.g., Wheeler & Fiske, 2005). The amygdala is generally associated with an initial affective fear response in reaction to the presentation of a potentially threatening stimulus in the environment (for review, see LeDoux, 2007; Balleine & Killcross, 2006). Furthermore, greater amygdala activation has been associated with greater levels of implicit racial bias (Phelps et al., 2000). This fear or anxiety response, be it conscious or unconscious, is inconsistent with most people’s social interaction goals. Therefore, self-regulation is necessary so that this initial response is not expressed (Stephan et al., 1999).

Social neuroscience research identifying neural mechanisms involved in implicit racial bias has contributed to a better understanding of the mechanisms that underlie cognition and behavior observed in the lab and also in the real world as well. Neural findings have suggested that there are at least three major interacting components involved in the expression and inhibition of implicit racial bias (Stanley, Banaji, & Phelps, 2008; Kubota, Banaji, & Phelps, 2012). As alluded to earlier, one of these regions is the amygdala, which is thought to be associated with an initial affective fear response in reaction to the presentation of a potentially threatening stimulus in the environment (i.e., presentation of a Black face; e.g., Wheeler & Fiske, 2005; for review, see Chekroud,
Everett, Bridge, & Hewstone, 2014). Research by Amodio, Harmon-Jones, and Devine (2003) has demonstrated that White Americans show fear-based, negative affective responses when Black faces are present. The authors instructed participants to view pictures of White and Black faces, each one subsequently followed by an acoustic startle probe (a burst of white noise). They found that, among White participants, Black faces potentiated the startle eye blink response, known to be a direct reflection of the central nucleus of the amygdala (for review, see Davis, 2006).

Another neural substrate commonly identified during functional magnetic resonance imaging (fMRI) investigations of racial bias is the dorsolateral prefrontal cortex (dLIPC). This region is thought to support the regulation of this initial threat-based fear reaction in order to generate more desired behavioral responses consistent with the knowledge that the stimulus is not actually threatening (e.g., Richeson et al., 2003; Cunningham et al., 2004; for review, see Kubota et al., 2012). However, anatomical projections suggest that the dLIPC and the amygdala generally do not interact directly (Bracht et al., 2009). A region known as the dorsal anterior cingulate cortex (dACC) is thought to generate a signal that a modified behavioral response is required to override and regulate the initial affective reaction originating from the amygdala that conflicts with egalitarian goals (Stanley et al., 2008; Kubota et al., 2012; Hajcak & Foti, 2008; Luu, Flaisch, & Tucker, 2000).

More generally, the dACC is thought to play a key role in an evaluative error and/or conflict detection system. The Error-Related Negativity (ERN) is an event-related potential (ERP) component that reflects activity in the dACC. The ERN waveform occurs immediately following the commission of errors (Gehring, Goss, Coles, Meyer, &
Donchin, 1993), and also when response conflict is present (Yeung, Botvinick, & Cohen, 2004). For this reason, it is thought to monitor for information that conflicts with the intended response (Braver, 2012; Holroyd & Coles, 2002). When conflicts are detected, it signals to other brain regions that more control is needed (Bartholow et al., 2005; Hajcak et al., 2008; Luu et al., 2000).

Several studies have examined neural responses to errors during a task designed to measure implicit racial bias. When errors are made that are indicative of unconsciously endorsing Black stereotypes, they elicit larger ERN amplitudes compared to errors without implications about one’s implicit attitudes (e.g., Amodio et al., 2004; Amodio, Devine, & Harmon-Jones, 2008; Bartholow, Henry, Lust, Saults, & Wood, 2012). These neural responses predict the degree to which bias can be controlled (Amodio et al., 2008; Bartholow et al., 2012). This suggests that the dACC plays a crucial role in the implementation of cognitive control.

However, the exact nature of this role during self-regulation appears to be more complex than a “cold,” non-emotional error detection system. This seems especially evident in the example described above, when the error represents the unintentional expression of implicit racial bias. When the attempt to exert cognitive control fails, it is met with negative affect, which in this context, is the subjective experience of compunction (Monteith, 1993). Research has shown this feeling of compunction is important and necessary for discouraging people from behaving in a prejudiced way (Monteith, 1993). Therefore, this error detection process appears to be much more than just a middleman in between the uninhibited, unconscious feeling and the regulatory processes that suppress it. Rather, the dACC is necessary for making errors salient and
aversive so that we feel motivated to behave differently in the future (Bartholow et al., 2012; Hajcak et al., 2008). In a more general context, recent theories posit that the ERN reflects error-related affect, or more specifically, the degree to which errors are distressing (Bartholow et al., 2012; Hajcak et al., 2008; Inzlicht et al., 2015; Spunt, Lieberman, Cohen, & Eisenberger, 2012). This subjective feeling of negative affect elicited by errors appears to be instrumental in engaging control (Bartholow et al, 2012; Aarts & Pourtois, 2013; Desender, Van Opstal, & Van den Bussche, 2014).

**Emotion Regulation**

Despite a growing volume of evidence establishing that the ERN does reflect affect, little is known about the nature of this affective component, and whether it is comparable to other emotions. If the ERN/dACC is indeed so closely tied to the affective experience of distress related to commission of errors, then perhaps it could be regulated or modified in the same way that emotions can be. The methods by which emotional states are intentionally modified has been studied extensively within psychology (e.g., Gross, 2007). Generally referred to as *emotion regulation*, it has been defined as “the processes by which individuals influence which emotions they have, when they have them, and how they experience and express these emotions” (Gross, 1998).

The predominant framework for conceptualizing differences between various ways of regulating emotions is the process model of emotion regulation (Gross, 1998). It distinguishes different types of regulation methods based on the timing of when the emotion modifying process is initiated in relation to the onset of the subjective experience of the emotion (Gross, 1998, 2001). One distinction the model makes is that some emotion regulation strategies are implemented before the emotional response has
completely materialized. These are referred to as antecedent-focused emotion regulation strategies. For example, one approach that has received a great deal of attention in the literature is cognitive reappraisal. It involves reframing the affective stimulus, or thinking about it from a different perspective, in order to alter the meaning attributed to the stimulus (e.g., Gross, 2001; for review, see Ochsner, Silvers, & Buhle, 2012).

Alternatively, there are also response-focused emotion regulation strategies, or those that don’t begin to unfold until after the emotion is already underway (Gross, 1998, 2001). One example is expressive suppression, which involves inhibiting the external behavioral signs of an emotional response (Gross, 2001, 2002; Gross & John, 2003; Gross & Levenson, 1993; Butler, Lee, & Gross, 2007; Goldin, McRae, Ramel, & Gross, 2008; Bebko, Fraconeri, Ochsner, and Chiao, 2011). While emotion regulation has been applied to a variety of different contexts, it has yet to be examined within the realm of controlling unintentional expression of racial bias. Currently, it remains unknown whether the distress associated with error commission is malleable via emotion regulation strategies, and if any of these affective changes would be reflected in the ERN.

**The Current Studies**

The purpose of the current studies was to investigate the impact of emotion regulation strategies on the inhibition and expression of implicit racial bias. Specifically, we sought to examine the effect of regulating the distress following the accidental expression of automatic bias on behavioral and neural manifestations of implicit racial attitudes.

In two EEG studies, participants engaged in emotion regulation strategies while performing a task measuring implicit racial bias (Weapons Identification Task; WIT;
Payne, 2001). During this task, participants categorize objects as either guns or tools as quickly as they can. Right before the object appears on the screen, they are primed with either a White or a Black face. When participants mistakenly misidentify a tool as a gun following the presentation of a Black face, this is thought to be indicative of implicit bias (e.g., Payne, 2001). While performing the WIT, they either suppressed or reappraised in a positive light the distress they experienced following an erroneous response. To the extent that committing bias-related errors is aversive and conflicts with current goals, attempting to regulate emotional responses to such errors might alter the expression and control of racial bias.

Previous research indicates that reappraisal is effectual at decreasing the experience of negative emotions (for review, see Ochsner et al., 2012), while expressive suppression is viewed as less effective and can even have detrimental long-term effects on mental and physical well-being (Gross, 1998; 2002; Gross & John, 2003), especially in American population samples with Western-European values (e.g., Butler, Lee, & Gross, 2007). While effectual emotion regulation strategies like reappraisal can serve as a positive coping mechanism, here we predicted that reappraising negative emotions caused by making errors would actually increase the behavioral expression of racial bias. Related to this decrease in performance, we predicted that there would also be a decrease in the influence of cognitive control processes (determined through process dissociation analysis; Payne, 2001). Furthermore, it was thought that this reduction in distress would be evident in the decreased amplitude of the ERN. If the ERN can be attenuated as the result of reappraisal, then this would further support the notion that the ERN does in fact reflect affect. Additionally, it would indicate that the error-related distress reflected in the
ERN could perhaps be modified online in the same way other emotional states (like fear) can be. The WIT contains two different types of errors, each with the potential to cause different emotional consequences: errors can be indicative of endorsing racial stereotypes (i.e., erroneously classifying a tool as a gun following the presentation of a Black face), or they can just be errors, devoid of any other meaning about one’s automatic biases (e.g., accidentally categorizing a gun as a tool after being primed with a White face). This allowed us to not only investigate the effect of emotion regulation strategies on error-related aversive feelings more generally, but also the effect on bias-related compunction. Specifically, we hypothesized that reappraising emotional responses to errors during the WIT would decrease the amplitude of the ERN and task performance in general, and also increase the expression of racially biased behavior (evident in the WIT). If reappraisal is successful at reducing both error-related distress in general, we expected to find an overall reduction in the amplitude of the ERN. If reappraisal is capable of reducing bias-related compunction, we would expect to observe a decrease in the divergence in ERN signal following errors indicative of bias (i.e., black-tool) and non-biased errors.

Consistent with previous research indicating that reappraisal is superior to suppress at altering emotional experiences (e.g., Gross & Levenson, 1993), we predicted that the behavioral effects described above would be smaller, or even reversed, during suppress compared to reappraise. Past research indicates that engaging in expressive suppression can sometimes reduce subjective experience of negative affect (Goldin et al., 2008), while the neural correlates of emotional response actually increased. This suggests that the neural correlate of error-related affect, the ERN, might actually increase during suppression. However, given that suppression is a response-focused regulation technique,
we would expect that it would not have any impact on the ERN, due to temporal limitations – the ERN occurs approximately 100 ms following erroneous responses, which may not be enough time to engage in suppression.

These hypotheses were tested in Study 1 using a within-subjects, three-factor, blocked design. EEG was recorded while participants completed the WIT in three separate blocks. During each block, participants were given different instructions regarding how they should cope with any emotions they might experience in response to errors. Participants were instructed to reappraise their errors in a way that reduced their negative affect during one block, and to suppress external signs of emotion during another block. Additionally, participants completed a control block (“attend”), during which they received the standard WIT instructions.

Study 1

Method

Participants. A total of 47 undergraduate volunteers (20 females) participated in the first study. Participants were all college-age students (mean age is 19) enrolled in the Introduction to Psychology course at the University of Missouri. The sample was 6% Hispanic, 9% Black, 9% Asian, 4% multiracial, and 72% White, which is representative of the university population. Participants were recruited through the subject pool through which they participated in exchange for course credits. Subjects were selected on the basis of eligibility and interest. Individuals determined to be eligible for EEG participation were allowed to take part in the study.

Measures.
**Behavioral Task.** Participants completed the Weapons Identification Task (WIT; Payne, 2001), which is a fast-paced, reaction time test. During the WIT, participants are primed with a Black or White face prior to categorizing an object as either a gun or a tool. The combination of primes (black, white) and targets (tool, gun) produces four different trial types. Prime and target combinations of black-gun and white-tool are considered *stereotype-congruent* trials, while black-tool and white-gun are *stereotype-incongruent* trials (Payne, 2001). It is assumed that knowledge or unconscious endorsement of racial stereotypes will cause people to be more inaccurate and slower to respond during the stereotype-incongruent trials.

Each trial begins with a black-and-white masking pattern that appears on the screen for 1000 ms. Then, a picture of either a White or a Black male’s face (prime stimulus) appeared for 200 ms, followed immediately by a picture of either a gun or a tool (target stimulus) displayed for 200 ms, after which a second masking pattern appeared onscreen for 300 ms. Participants were given 500ms after the onset of the target stimulus to respond by pressing one of two buttons indicating whether the image was a gun or a tool. The response deadline coincided with the offset of the second masking pattern. If a response occurred after the deadline, it elicited the feedback message (“Too slow!”), alerting the subject that their response was not fast enough. During the inter-trial interval (ITI), a “+” is displayed on the screen for either 800, 1000, or 1200 ms.

Black-and-white photographic images were used for both prime and target stimuli. Eight images of male faces cropped to display only the midsection of the face (from the lips to eyebrows) were used as the primes (four black and four white). The target stimuli consisted of four pictures of different tools (e.g. a wrench), and four
pictures of different handguns, all of which were depicted against a solid white or light grey background.

**Affect.** The Positive and Negative Affect Scale (PANAS; Watson, Clark, & Tellegen, 1988) was used to assess current levels of positive and negative feelings. On a scale of 1 (*very slightly or not at all*) to 5 (*extremely*), participants rated the extent to which they presently felt each of 10 different positive (e.g., content) and 10 negative affective states (e.g., bored). Additionally, the State-Trait Anxiety Inventory (STAI; Spielberger, 1983) was also completed at baseline to measure current state (STAI-S) and more general trait (STAI-T) levels of anxiety. Participants responded with how much they were experiencing (or experience in general) each of 20 emotional descriptions (e.g., “I feel calm” or “I feel nervous and restless”) on a scale of 0 (*not at all/almost never*) to 3 (*very much so/almost always*). All three of these questionnaires were administered at baseline before beginning the task, and the PANAS and STAI-S were also given at the end of each block to track changes in self-reported emotions as the experiment progressed.

**Manipulation check.** Participants completed an in-house Post-Experiment Questionnaire (PEQ) to address questions regarding the subjective experience of participating in the study and to ensure that participants followed the instructions properly. Questions included items such as “Were you successful at suppressing your feelings about making errors when you were asked to do so?” and also whether or not they preferred one strategy over the other, and if they found one to be easier.

**Emotion regulation style.** In order to assess individual differences in tendencies to use reappraisal and suppression when coping with emotions in everyday life, they were
given the Emotion Regulation Questionnaire (ERQ; Gross & John, 2003). Subjects indicated their agreement or disagreement with each of 10 statements (e.g., “I keep my emotions to myself” or “When I want to feel more positive emotions, I change the way I’m thinking about the situation I’m in”) on a scale from 1 (strongly disagree) to 7 (strongly agree).

**Social desirability.** An abbreviated version of the Marlow-Crowne Social Desirability Scale (Marlow & Crowne, 1960) was administered to assess a participant’s tendency to give responses consistent with how they think the experimenter wants them to respond. This questionnaire is used widely, especially in studies where subjects are asked to engage in emotion regulation strategies, such as reappraisal (see Ochsner, Bunge, Gross, and Gabrieli, 2002). Participants indicated whether they found a series of 33 items (e.g., “I’m always willing to admit it when I make a mistake” and “I have never deliberately said something that hurt someone’s feelings”) to be true or false.

**Racism.** Explicit racial bias against Black people was measured with two different scales. The Modern Racism Scale (MRS; McConahay, 1986) is a six-item questionnaire that assesses racial attitudes toward Black people. Participants specified their agreement with each statement (e.g., “Over the past few years, Blacks have gotten more economically than they deserve”) on a scale of 1 (strongly disagree) to 5 (strongly agree). The Motivation to Respond without Prejudice Scale (MRPS; Plant & Devine, 1998) was used to measure the extent to which one’s motivation to behave in an unbiased way is driven by internal and external pressures (two separate scales). Participants were asked to rate agreement with ten statements on a scale of 1 (strongly disagree) to 9 (strongly agree). Half of these items belonged to the internal motivation scale (e.g., “I
attempt to act in nonprejudiced ways toward Black people because it is personally important to me”) and the other half was meant to capture external motivation (e.g., “because of today’s PC (politically correct) standards I try to appear nonprejudiced toward Black people”).

**Electroencephalogram.** The electroencephalogram (EEG) was recorded using an array of 24 silver/silver-chloride electrodes fixed in a stretch-lycra cap (ElectroCap, Eaton, OH) placed on the scalp in standard locations (American Encephalographic Society, 1994). Additional bipolar electrodes were placed on the face to measure the electrooculogram. Conductive gel was inserted in each electrode to reduce impedance between the electrodes and scalp, which was kept below 8 kΩ at all locations. EEG signals were amplified with Synamps2 amplifiers (Compumedics-Neuroscan, Charlotte, NC), sampled at 500 Hz, filtered online at 0.01-40 Hz, and referenced online to the right mastoid. Offline, an average-mastoid reference was derived, and ocular artifacts were removed using a regression-based procedure (Semlitsch, Anderer, Schuster, & Presslich, 1986). A band-pass filter of 1 to 15 Hz was applied prior to deriving Response-locked epochs of 1,000 ms (400 ms pre-response baseline). After signal artifacts were rejected manually, an average signal was obtained for each electrode for each participant per cell.

**Emotion regulation manipulation.** The experimental manipulation consists of a within-subjects comparison of two commonly studied emotion regulation strategies, reappraisal and suppression. During each block, participants received one of three different instructions regarding how they should think about any errors they make while performing the WIT. For the first block (attend), subjects were simply given the standard WIT instructions. The attend block was meant to be a baseline measure of WIT
performance without the application of any emotion regulation strategy, and always occurred first in order to prevent interference from previous instructions. The emotion regulation instructions were given during the second and third blocks (order counterbalanced across participants). During the suppression block, participants were told to not express any externally visible, behavioral signs of emotions whenever they made errors (Gross & Levenson, 1993; Goldin et al., 2008; Bebko et al., 2011). For the reappraisal block, they were asked to “regulate” their emotions by thinking about their errors in such a way that made them feel less negative about them (e.g., by telling themselves “everyone makes mistakes” following each error; Ochsner et al., 2002).

**Procedure.** Once they have arrived at the Social Cognitive Neuroscience Lab on the University of Missouri campus, all participants were given informed consent. Afterwards, the EEG cap was applied to the participant’s scalp. Prior to receiving task instructions, participants completed a set of brief questionnaires, the first of which inquired about demographics (age, sex, race, ethnicity and nationality). Additionally, they completed the STAI-T, STAI-S, and PANAS to assess their level of anxiety in general (STAI-T), and also gauge their baseline emotional state.

Prior to beginning the WIT, participants were made aware that the computer task they would be performing was designed to measure individual differences in unconscious racial bias. This information was intended to increase participants’ incentive to respond accurately during the task and to help ensure that they experienced feelings related to compunction when they made errors. After receiving instructions for the WIT and being given an opportunity to practice (8 trials), they performed the task while event-related brain potentials (ERPs) were recorded using electrodes affixed to the scalp. Each of the
three blocks consisted of 192 trials, which included 48 trials per prime-target pair. Each block took approximately ten minutes to complete, and participants were given a self-timed break at the half-way point.

At the end of each block, participants were asked to complete two brief questionnaires (PANAS and STAI-S), which took no more than five minutes. Before beginning the second block, the experimenter returned to instruct participants in one of the two emotion regulation strategies (suppress or reappraise). The same procedure was repeated for the third block, where participants received the instructions for the remaining emotion regulation strategy. Afterwards, the cap and the electrodes were removed, and the subjects were given a chance to get cleaned up before completing a final set of five questionnaires (PEQ, ERQ, MC-SDS, MRS, and MRPS), and then being debriefed.

**Design.** This experiment examined the accidental expression of automatic bias, and the impact of regulating the distress associated with these errors. We compared performance while engaging in different emotion regulation strategies during trials that are stereotype-incongruent (black-tool and white-gun) and trials that are stereotype-congruent (black-gun and white-tool). This produced a 3 (Block Instructions; attend, suppress, reappraise) x 2 (Prime; black, white) x 2 (Target; tool, gun) within-subjects design.

**Results**

**Behavioral data.**

**Analytic approach.** Although all participants were included when analyses permitted, one of the three blocks was excluded for two individuals. One participant failed to adhere to instructions during the last block (suppress) by responding with only
one of the two buttons for all trials. Another participant mistook the power drill tool for a
gun during the first block (attend). Accuracy data was analyzed by calculating the
proportion of accurate trials (excluding missed trials) for each subject. These values were
subsequently used to obtain measures of automatic and controlled processes influencing
responses for each subject via an analysis known as process dissociation procedure (PDP;
Payne, 2001). This technique has been previously used to disentangle the automatic and
controlled processes contributing to performance during the WIT. This method was
originally developed by Jacoby (1991) for the purpose of separating automatic from
controlled processes influencing memory recall, and was later applied to the WIT by
Payne (2001). Since then, the PDP has been used repeatedly in conjunction with the WIT
(e.g., Payne, 2005; Huntsinger, Sinclair, & Clore, 2009; Bartholow et al., 2012; Amodio
et al., 2004, 2008; Stewart & Payne, 2008; Huntsinger, Sinclair, Dunn, & Clore, 2010;
Govorun & Payne, 2006; Huntsinger, 2013). For an in-depth description of how the
values for controlled and automatic processes are derived from accuracy rates, see Payne
(2001, 2005). PDP scores for controlled and automatic processes were calculated for each
subject separately within each condition.

**Accuracy during the WIT.** The proportion of accurate trials per subject were
submitted to a 3 (Block Instructions: attend, reappraise, suppress) x 2 (Prime: black,
white) x 2 (Target: gun, tool) repeated measures ANOVA,\(^1\) which revealed a significant
Prime x Target interaction \(F(1, 44) = 80.51, \ p < .001, \ \eta_p^2 = 0.65; \) Figure 1a). When the
\(^1\) Note that the results did not differ when transformed accuracy rates (arcsine of the
square root) were submitted to the same analyses. For the sake of simpler interpretation
of results, the non-transformed accuracy results are presented here.
prime was a Black face, participants were less accurate at identifying tools ($t(44) = -7.39, p < .001$) and more accurate at identifying guns ($t(44) = 9.10, p < .001$), compared to when the prime was a White face. Likewise, accuracy was greater during gun than tool trials when primed with a Black face ($t(44) = 6.33, p < .001$); however, this pattern was reversed when primed with a White face ($t(44) = -3.44, p = .001$). This pattern is evidence of implicit bias and is typical for the WIT (e.g., Payne, 2005; Bartholow et al., 2012; Amodio et al., 2004, 2008). Because this finding is standard (and is not a direct test of our hypotheses), this Prime x Target interaction (along with qualifying main effects) will only be mentioned briefly in future analyses without further elaboration. The ANOVA also yielded a main effect for block instructions ($F(2, 88) = 16.64, p < .001, \eta^2_p = 0.27$). Planned pairwise comparisons revealed that accuracy during attend ($M = .71$) was greater than reappraise ($M = .67, t(44) = 3.23, p = .002$), and that both attend and reappraise were greater than suppress ($M = .65, t(45) = 5.04, p < .001$; and $t(45) = -2.99, p = .004$, respectively; Figure 1a).

Furthermore, there was a significant three-way interaction of Block Instructions x Prime x Target ($F(2, 88) = 4.21, p = .018, \eta^2_p = 0.09$). Inspection of the means in Figure 1a suggested that this interaction might be driven by differential effects of prime race across instruction conditions, especially for gun targets. To investigate this possibility, separate 3 (Block Instructions) x 2 (Prime) repeated measures ANOVA were conducted.

---

2 All participants were included in analyses when possible. Since one block has been excluded for two different subjects, they are included in some of the t-tests, depending on the comparison. This is why there is some variation in degrees of freedom associated with different t-tests.
for the tool and gun target trials. The ANOVA on tool trials showed main effects of block instructions ($F(2, 88) = 5.73, p = .005, \eta^2_p = 0.12$) and prime ($F(1, 44) = 54.62, p < .001, \eta^2_p = 0.56$), but no interaction. In contrast, the ANOVA on gun trials showed a significant interaction of Block Instructions x Prime ($F(2, 88) = 5.13, p = .008, \eta^2_p = 0.10$), which seems to be driven by the exaggerated difference between black- and white-gun trials evident in the emotion regulation blocks (Figure 1a; Table 1). This divergence is caused by the plunge in accuracy during white-gun trials for reappraise ($t(45) = 4.29, p < .001$) and suppress ($t(46) = 4.22, p < .001$) relative to attend. Additionally, the Prime x Target interaction was tested separately within each block. We found that the size of the interaction effect is larger in suppress ($F(1, 46) = 57.67, p < .001, \eta^2_p = 0.56$) than attend ($F(1, 45) = 28.17, p < .001, \eta^2_p = 0.39$), and even larger in reappraise ($F(1, 45) = 80.60, p < .001, \eta^2_p = 0.64$).³

To acquire a quantified measure of racial bias based on WIT performance, we subtracted the accuracy rate of black-tool trials from white-tool trials to create a difference score for each participant. Amodio and colleagues (2004) have argued that the race priming effect evident during tool trials serves as an indication of racial bias, in that a bias-free response set would be expected to show no differences on tool trials regardless of prime condition. Planned comparisons between blocks suggest that this difference is significantly greater during reappraise ($\Delta = .15$) compared to suppress ($\Delta = .11, t(45) = 2.08, p = .043$), and although it did not reach significance, there was a trending difference

³ When the same analysis was applied to the subset of White participants ($n = 34$), the same effects and interactions found when using data from the full sample were reproduced.
between attend (Δ = .11) and reappraise (t(44) = 1.77, p = .084). Thus, it seems that reappraising errors caused participants to exhibit a more biased response pattern compared to suppress, and perhaps also compared to attend.

To summarize, participants were less accurate while performing suppress across all WIT conditions relative to attend; this difference was especially pronounced during white-gun trials. During reappraise, accuracy within each WIT trial type did not differ from attend, except for a drastic decrease during white-gun trials. As evidenced by the larger discrepancy between accuracy on black-tool and white-tool trials, participants exhibited more racial bias during reappraise compared to suppress (and marginally more compared to attend).

**Response time during the WIT.** Response time (RT) in milliseconds during correct trials was analyzed using a 3 (Block Instructions) x 2 (Prime) x 2 (Target) repeated measures ANOVA. A main effect for target (F(1, 44) = 25.10, p < .001, ηp² = 0.36) indicates that participants were slower to respond during tool (M = 343.63 ms) compared to gun (M = 332.10 ms) trials. This effect was qualified by a significant interaction of Prime x Target (F(1, 44) = 43.83, p < .001, ηp² = 0.50). Examination of the means in Table 1 shows that RTs were slower during tool compared to gun trials when primed with a Black face (t(45) = -7.96, p < .001; Table 1), whereas tool and gun target RTs following White face primes did not differ (t < 1, p > .8). There was a main effect for block instructions (F(2, 88) = 25.56, p < .001, ηp² = 0.37), indicating that participants took longer to respond during attend (Table 1) compared to both suppress (Table 1; t(45) 4 Note that because log-transformed RTs yielded similar results to untransformed RTs, we will instead present the untransformed RTs in order to ease interpretation.
= 6.04, \( p < .001 \) ) and reappraise (Table 1; \( t(44) = 4.69, p < .001 \)) . They were also slower during reappraise than during suppress (Table 1; \( t(45) = -3.12, p = .003 \)).

**Automatic and controlled processes.** PDP scores for controlled and automatic processes were analyzed separately in 3 (Block Instructions) x 2 (Prime) repeated measures ANOVAs (Figure 1b). Analysis of automatic processes revealed a main effect for prime (\( F(1, 45) = 92.93, p < .001, \eta_p^2 = 0.67 \)), meaning that the influence of automatic processes on responses was greater during black prime trials compared to white (Figure 1b; note that this finding is typical for the WIT). Additionally, there was a significant main effect of block instructions (\( F(1, 45) = 8.88, p = .005, \eta_p^2 = 0.17 \) ). Simple effects analyses indicated that automatic processes were greater during attend compared to both suppress (\( \Delta = .03, t(45) = 2.20, p = .033 \) ) and reappraise (\( \Delta = .05, t(45) = 2.98, p = .005 \); Figure 1b). These main effects qualified the significant two-way interaction of Block Instructions x Prime (\( F(1, 45) = 6.70, p = .013, \eta_p^2 = 0.13 \) ). Pairwise comparisons indicate that automatic processes influencing responses to black primes did not fluctuate across blocks (ts < 1, ps > .3). However, for white primes, automatic processes were reduced during reappraise compared to both attend (\( t(45) = 3.98, p < .001 \) ) and suppress (\( t(45) = 2.03, p = .049 \); Figure 1b).

In the analysis of controlled processes, only the effect of block instructions was significant (\( F(1, 45) = 9.81, p = .003, \eta_p^2 = 0.18 \) ). Regardless of prime type, participants implemented more control overall during attend (\( M = .43 \) ) compared to both emotion regulation strategies (reappraise: \( t(45) = 3.13, p = .003 \); suppress: \( t(45) = 5.20, p < .001 \) ), and also during reappraise (\( M = .36 \) ) compared to suppress (\( M = .30; t(45) = 2.99, p = .004 \) ). This suggests that both emotion regulation strategies reduced the level of control.
exhibited during the WIT, with a greater reduction occurring while participants were engaged in suppress (Figure 1b).

**ERP data.**

*Analytic approach.* To account for the multilevel nature of ERP data, we analyzed the ERN using a multilevel model (MLM) with restricted maximum likelihood estimation method. MLM has been preferred over repeated measures ANOVA (e.g., Stroup, 2014) in more recent years within the field of psychology, especially when modeling psychophysiological variables (e.g., Bailey et al., 2014; Hilgard, Weinberg, Proudfit, & Bartholow, 2014; Tritt, Page-Gould, Peterson, & Inzlicht, 2014). Additionally, unlike repeated measures ANOVA, MLM is robust to missing data and allows for different numbers of trials across conditions, thus preventing us from excluding participants on the basis of missing data. Degrees of freedom were calculated using the containment method. A variance components covariance structure was used to estimate random intercept coefficients for participants and electrodes nested within participants. Planned comparisons were conducted by comparing the estimated marginal population means via paired t-tests (calculated with the LSMEANS /DIFF statement in SAS). We analyzed the quantified average for each ERP component using a 3 (Block Instructions: Attend, Reappraise, and Suppress) x 2 (Prime: Black, White) x 2 (Target: Gun, Tool) MLM with random intercepts for subject and electrode channel nested within subject. The means reported represent the estimated marginal means within the specified model, and are measured in units of microvolts (µV).

Averages for each cell within the 3 x 2 x 2 design included only the participants with valid (artifact-free) EEG data for at least five error trials. Importantly, this means
that participants were not excluded on a listwise basis; they were included in analyses if they had at least five valid error trials within at least one cell. ERN waveforms were quantified per subject within each cell (again with a minimum of five errors) by measuring the average signal over the interval of 30 to 110 ms (determined by visual inspection) following the response. Analyses were conducted using the 3 x 3 mid-frontal electrode sites centered on the location at which the peak ERN occurred (FCz). Thus, the analyses included three frontal midline channels (Fz, FCz, and Cz), three lateral locations left of the midline (F1, FC1, and C1), and their right-hemisphere homologues (Figure 2a).

It is important to note that some researchers have argued for the importance of looking at the ERN in relation to the negativity in the signal also that happens following a correct response, i.e., the correct-response negativity, or CRN. By subtracting the CRN from the signal, one is presumably left with the negativity that specifically results from committing errors, thereby controlling for any negativity that may have occurred as the result of performing the task in general (Luck, 2005). However, examination of the ERN-CRN difference waveform was deemed to be not appropriate for this paradigm. During this task, participants were not regulating emotional responses following correct trials. Therefore, this procedure would not be applicable here, because the tasks performed during correct and error trials were slightly different (e.g., during the reappraise block, when following errors, they were reappraising, while after correct responses, they were not reappraising).

**ERN.** The quantified ERN data were analyzed using a 3 (Block Instructions) x 2 (Prime) x 2 (Target) MLM with random intercepts for subject and nine electrode channels nested within each subject. The results revealed main effects for prime ($F(1,$
4165) = 24.77, \( p < .001 \) and target (\( F(1, 4165) = 16.52, p < .001 \)), which were qualified by a Prime x Target interaction (\( F(1, 4165) = 25.75, p < .001 \)). These results are explained by the enhanced ERN amplitude during black-tool trials (\( M = -1.81 \mu V \) relative to all other WIT trial types when collapsed across blocks (\( Ms \) range from -0.86 to -0.95; Figure 2c). There is also a main effect for block instructions (\( F(2, 4165) = 28.85, p < .001 \)), indicating that the ERN amplitude was greater during attend (\( M = -1.60 \)) compared to both suppress (\( M = -1.09; t(4165) = -7.58, p < .001 \)) and reappraise (\( M = -0.73; t(4165) = -4.27, p < .001 \); Figure 2c). Average ERN amplitudes are also greater during suppress relative to reappraise (\( t(4165) = 3.13, p = .002 \); Figure 2c). The MLM analysis also revealed a significant interaction of Block Instructions x Prime (\( F(2, 4165) = 9.43, p < .001 \)). There is no difference between ERNs (averaged across target) elicited during black and white prime trials during the attend block (\( |t| < 1, p > .8 \); Figure 2b). However, during both emotion regulation blocks, the ERN signal during white-prime trials was attenuated, thereby increasing the divergence between black and white prime responses (suppress: \( t(4165) = 2.83, p = .005 \), reappraise: \( t(4165) = 6.06, p < .001 \); Figure 2b).5

**Study 1 Discussion**

The purpose of this first study was to demonstrate in a within-subjects design that emotion regulation strategies (cognitive reappraisal and expressive suppression) could be used to reduce the negative affect that typically accompanies behaving in a racially biased way. First, we hypothesized that successful emotion regulation of error-related

---

5 This same model was tested using only the white participants (\( n = 34 \)). In spite of the smaller sample size, this analysis still yielded the exact same results as the analysis of the full sample.
(aversive) negative affect would reduce motivation to change behavior on future trials, and result in participants making more errors overall. Furthermore, we expected that this reduction would be greater during reappraise compared to suppress, consistent with previous research indicating that reappraisal is superior to suppress at altering emotional (subjective) experiences (e.g., Gross & Levenson, 1993). In partial support of our hypothesis, we found that reappraisal led to decreased accuracy overall (across all WIT conditions) relative to attend. However, contrary to expectations, suppress reduced accuracy more than both reappraise and attend, meaning that suppress may have actually been more effective than reappraise at reducing general aversive responses to errors.

Our second hypothesis was that successfully regulating negative emotions caused by making errors would not only increase error rates generally, but in a way that is indicative of an increase in the behavioral expression of racial bias during the WIT. As indications of increased racial bias, we expected to find the following (in comparison to other blocks): (1) a more robust prime x target accuracy interaction (i.e., the typical WIT response pattern; measured by effect size), (2) a greater discrepancy between black-tool and white-tool accuracy, and (3) automatic processes having greater influence over response outcomes (measured with PDP). In Study 1, we found that the interaction of prime and target was more robust in suppress than attend, and even larger in reappraise. However, this seems to be primarily driven by the exaggerated decrease in accuracy observed for white-gun trials during both suppress and reappraise blocks. Automatic processes influencing responses to black primes did not fluctuate across blocks, whereas for white primes, automatic processes were reduced during reappraise compared to both attend and suppress. Taken together, these results suggest that both emotion regulation
strategies reduce biased responses more generally (i.e., white bias), although this is difficult to interpret. Addressing the question of Black implicit racial bias specifically, the greater error rate during black- relative to white-tool trials indicated that participants exhibited more (Black) racial bias during reappraise compared to suppress (and marginally more compared to attend). More relevant to the overarching goals of current study, these results that reappraising errors caused participants to exhibit more (Black) racial bias compared to suppress, and marginally compared to attend.

Our third hypothesis was that this decrease in accuracy and enhanced expression of racial bias could be explained by an overall decrease in regulatory control. That is, if participants found errors to be less aversive, they would be less motivated to control bias and errors, and therefore, they would exhibit less control. If participants were implementing less control, and responding less carefully, we might expect a decrease in RTs (e.g., similar to the rationale for interpreting post-error slowing). Additionally, we predicted that there would also be a decrease in the influence of cognitive control processes on responses (measured with PDP). RT results indicated that participants took less time to respond during both suppress and reappraise compared to attend, and that responses during suppress were faster than reappraise. The PDP analysis showed that both emotion regulation strategies reduced the level of control exhibited during the WIT, with a greater reduction occurring while participants were engaged in suppress. Consistent with our hypotheses, these results demonstrate that reappraising errors during the WIT effectively reduced cognitive control engagement, thus leading to the decrease in overall accuracy and increased expression of racial bias described above. However,
contrary to expectations, this evidence indicates that expressively suppressing emotional responses decreased cognitive control more so than reappraisal.

For our fourth hypothesis, we thought that this regulation in error-related distress (which presumably caused the subsequent decrease in the implementation of control) would be evidenced by decreased ERN amplitude. Again, we thought that the ERN amplitude would be more attenuated (relative to attend) during reappraise than suppress. More specifically, we predicted that during reappraise, the ERN amplitude would be reduced overall relative to both suppress and attend, meaning that reappraise had successfully reduced the aversive feelings associated with making errors (in general) during the task. A second prediction following from this hypothesis was that reappraisal would more effectively reduce feelings of compunction that arise only after committing errors (i.e., during black-tool trials) indicating that one unconsciously endorses Black stereotypes. Typically during the WIT paradigm, the ERN is enhanced during black-tool trials (e.g., Bartholow et al., 2012). If reappraisal is successful at reducing all aversive feelings associated with error commission (including compunction following the commission of bias-related errors), then we would expect that the ERN during black-tool trials will be less enhanced relative to other trial types (within a given block).

Consistent with this hypothesis, a main effect of block instructions indicated that the overall ERN amplitude was reduced during suppress compared to attend, and even more so during reappraise, such that it was less than both attend and suppress. This suggests that while both emotion regulation strategies were successful at reducing error-related distress, reappraise was more effective than suppress. This is at least partially consistent with previous research comparing the efficacy of reappraise and suppress (e.g.,
Goldin et al., 2008). We also found an interaction of block instructions and prime, whereby the ERN amplitude during white-prime trials was more attenuated (less negative) during both emotion regulation blocks. This caused the divergence between black and white primes observed in the ERN to actually increase during both emotion regulation blocks compared to the attend block, which is the opposite of what we predicted. Additionally, we found that during black-tool trials in particular, the ERN amplitude remained elevated across all three block instructions, while the other WIT conditions decreased (Figure 2c). This suggests that attempts to regulate feelings of distress from erroneous responses were successful some of the time, but not when errors were indicative of implicit stereotype endorsement. Taken together, the ERN results from Study 1 suggest that both regulation strategies (especially reappraisal) are effective at reducing error-related distress more generally, but they are not as effective when this distress is coupled with the feeling of compunction following a biased error.

One aspect of these results that violated our expectations was that suppress consistently out-performed reappraisal as the more effective emotion regulation strategy for reducing distress from error commission across multiple converging measures. On the whole, this is inconsistent with previous research rendering suppress as ineffective (e.g., Gross & Levenson, 1993). One possible explanation for this finding is that any benefit obtained by engaging in expressive suppression may be due to attention shifting from the distress caused by making errors to the task of suppressing outward signs of emotions (Goldin et al., 2008). That is, perhaps participants were distracted from the aversive consequences of making errors, and this allowed the strategy to be effective. Another possibility is that suppression is simply well-suited for this paradigm. However, this latter
explanation is doubtful, given that there is nothing special about this paradigm that would lead to results that contradict most of the previous findings about suppression.

Among the limitations of Study 1, there are other alternative hypotheses that could explain differences in both emotion regulation blocks that the present design fails to eliminate. One major concern is that attend always occurred during the first block. As such, the observed reductions in accuracy, control, and ERN amplitude during the emotion regulation blocks could be merely due to fatigue, rather than a change in affective state. Another alternative hypothesis is that increased error rates are actually due to increased cognitive demand instead of diminished error-induced distress. This is based on the assumption that performing an emotion regulation strategy in addition to the WIT would be more difficult than performing the WIT by itself. This increased difficulty could be more taxing on cognitive resources, thereby leading to worse performance. Additionally, one might argue that during Study 1, a carry-over effect may have affected the third block. This means that emotion regulation instructions from the preceding block might have affected ability to perform the instructions issued for the subsequent block. It may have been difficult for participants to discontinue instructions for one strategy and commence utilizing a different strategy applied to the same task.

**Study 2**

In order to address some of these limitations identified in relation to the within-subjects manipulation in the original study, we conducted a second, follow-up study to incorporate a between-subjects comparison of the different block instructions. Since all participants in Study 1 had already completed the experiment with attend as their first block, we would count them as our attend group. (This reasoning was based on the
assumption that subsequent blocks would have no effect on performance during the 1st block.) In Study 2, participants were randomly assigned to one of two groups, which dictated whether they performed suppress or reappraise during the first block. All participants only completed two blocks, the second of which was always attend regardless of group assignment. As such, each participant was asked to perform only one of the emotion regulation strategies. The rationale for this modification was to reduce the possibility of a carry-over effect by ensuring that performance during one of the emotion regulation blocks would not be affected by conflicting instructions from a preceding block.

This design would allow us to make a between-subjects comparison of block instructions by comparing the first block across all subjects from Studies 1 (attend) and 2 (suppress and reappraise). This design would eliminate fatigue as a valid explanation for the results found in Study 1. Additionally, this analysis would render the carry-over effect issue obsolete. Furthermore, this would also allow us to probe whether the order in which the block instructions are administered is capable of affecting the results. Since participants in Study 1 always completed the attend block first, followed by one of the regulation blocks, we intended to compare these first two blocks with the two blocks completed by subjects in Study 2. This analysis would perhaps help to disentangle the cognitive load effect, while simultaneously examining whether fatigue was contributing to the effects found in Study 1.

Method

Participants. A priori, it was decided that there would be 30 participants in each of two groups. Data were collected from 67 participants in total, two of which were
excluded completely due to a failure to understand and follow instructions. The first block (reappraise) was excluded for one participant for the same reason, and the second block (attend) was excluded for six participants due to difficulty with discontinuing the regulation strategy (performed during the first block). Unfortunately, the EEG data for four valid subjects was lost following equipment malfunction. Partial data from three of those four has been retained (i.e., the data processed for analyses of errors), and will be included when it is possible. The final sample includes 65 participants total (29 females; mean age = 18.58), only 60 of whom exists valid behavioral data from the first block accompanied by utilizable EEG data (30 in each group; partial data from the other five participants have been included in analyses when possible). Based on the full sample of 65, approximately 11% of participants were Black, 8% Asian, 2% Hispanic, 5% multiracial, and the remaining 75% were White.

**Emotion regulation manipulation.** Participants were randomly assigned to one of two emotion regulation groups, that determined whether they completed suppress or reappraise as their first block. Participants in the current study were only asked to complete two blocks, the second of which was always attend. The specific emotion regulation strategy instructions remained unaltered. For the attend block, subjects were instructed to simply react how they normally would whenever they made an error (Ochsner et al., 2002).

**Procedure.** Most of the procedures remained unchanged from Study 1. During Study 2, before beginning the first block, participants were given sixteen practice trials instead of eight. The purpose for this was to provide participants with ample opportunity
to acquaint themselves with the WIT before being given the emotion regulation instructions for the first block.

To prevent against any carryover effects from the emotion regulation strategy instructions given during the first block, participants were given a filler task prior to beginning the second block. This was administered immediately following the completion of the post-block questionnaires for the first block, and right before receiving instructions for the second block. During the filler task, they were given an abstract, geometric design and asked to color with crayons for 5 minutes. Participants were told that the purpose of this activity was to “help break up the first and second blocks, and to clear their mind” before being given different instructions for the second block. Then, the experimenter returned to remove the coloring supplies and give the participants the instructions for attend. All other measures and other procedures did not differ from Study 1. Participants still completed the same post-block and post-experiment questionnaires as before.

**Results**

**Analytic approach.** In order to examine the difference between the emotion regulation strategies in a between-subjects comparison, performance and electrocortical activity during the first block for subjects across Studies 1 and 2 were analyzed. As mentioned previously, participants in Study 1 comprised the attend group, and participants in Study 2 were either in the reappraise or suppress group. For the following analyses, there were 46 subjects in the attend group, 30 in the suppress group, and 34 in the reappraise group. The same analytic approaches and quantification practices for ERPs described previously were applied here as well for each analysis. Both ERP and
behavioral data were analyzed using the same models described previously, except that block instructions was now a between-subjects predictor.

**Behavioral data.**

**Accuracy.** The accuracy data was analyzed in a 3 (Block Instructions) x 2 (Prime) x 2 (Target) mixed factorial ANOVA, with the latter two as within-subject factors. This yielded a main effect for target ($F(1, 107) = 10.84, p = .001, \eta_p^2 = 0.09$), suggesting that accuracy rates were greater during gun compared to tool trials (see Table 2; Figure 3a). These results returned the typical WIT pattern of the two-way Prime x Target interaction discussed in Study 1 (Table 2; $F(1, 107) = 76.45, p < .001, \eta_p^2 = 0.42$). Contrary to expectations, there was no significant effect for block instructions ($p = .163$).

**Response time.** RTs were also analyzed using a 3 (Block Instructions) x 2 (Prime) x 2 (Target) mixed factorial ANOVA. Similar to Study 1, there was a main effect for target ($F(1, 107) = 28.29, p < .001, \eta_p^2 = 0.21$) and an interaction of Prime x Target ($F(1, 107) = 58.57, p < .001, \eta_p^2 = 0.35$), suggesting that participants were slower to respond during tool than gun trials after being primed with a Black face ($t(109) = 8.05, p < .001$; Table 2), but not a White face ($|t| < 1, p > .8$). Additionally, the three-way interaction of Prime x Target x Instructions was marginally significant ($F(2, 107) = 2.96, p = .056, \eta_p^2$).

---

As before, the same results were obtained when this analysis was conducted using the transformed accuracy rates (i.e., arcsine of the square root). The same model (with raw accuracy data) was also tested using a reduced sample of only the white participants ($ns = 34, 27, and 21$ for attend, reappraise, and suppress, respectively). As reported in Study 1, the effects that were significant did not change.

Log-transformed RTs submitted to this analysis obtained the exact same effects.
= 0.05). By examining the means, it seems that the Prime x Target interaction differs across block instructions (Table 2); this was tested by conducting follow-up 2 (Prime) x 2 (Target) ANOVAs separately at each level of block instructions. Although the two-way interaction is significant within each block instructions, this effect size ($\eta^2_p$) is more robust in suppress ($F(1, 29) = 35.90, p < .001, \eta^2_p = 0.55$) relative to both reappraise ($F(1, 32) = 11.34, p = .002, \eta^2_p = 0.26$) and attend ($F(1, 46) = 17.48, p < .001, \eta^2_p = 0.28$). It appears that this is the primary cause of this marginal three-way interaction.

**Automatic and controlled processes.** PDP scores for automatic and controlled processes were analyzed separately using 3 (Block Instructions) x 2 (Prime) mixed factorial ANOVAs. Analysis of automatic processes yielded a significant main effect of prime ($F(1, 107) = 50.73, p < .001, \eta^2_p = 0.32$), suggesting that automatic processes contributed more to responses following black primes than white primes (see Figure 3b). No significant effects were found in the controlled processes analysis. Unlike the within-subjects comparison in Study 1, in the present between-subjects analysis, block instructions had no effect on controlled or automatic processes.

**ERP data.**

**ERN.** The 3 (Block Instructions) x 2 (Prime) x 2 (Target) MLM analysis of the ERN revealed main effects for both prime ($F(1, 2406) = 32.32, p < .001$) and target ($F(1, 2406) = 13.29, p < .001$), indicating that ERN amplitudes were greater for Black ($M = -2.43$) compared to White ($M = -1.75$) faces, and also for tools ($M = -2.31$) relative to guns ($M = -1.87$; Figure 2d). Additionally, there was a two-way interaction of Instructions x Prime ($F(1, 2406) = 10.91, p < .001$); like Study 1, we observed no difference between black ($M = -1.58$) and white ($M = -1.61$) primes during attend ($t < 1, p > .8$), whereas
black primes elicit a more negative ERN than white during both suppress ($M_s = -2.15$ and $-1.26$ for black and white, respectively; $t(2406) = -3.97, p < .001$) and reappraise ($M_s = -3.57$ and $-2.39$ for black and white, respectively; $t(2406) = -3.97, p < .001$; Figure 2d). Additionally, the reappraisal instructions generated enhanced ERN amplitude during black prime trials compared to attend ($t(2406) = -2.38, p = .017$). A significant interaction of Instructions x Target ($F(1, 2406) = 5.89, p = .003$) indicates that across primes, the ERNs appear to be greater during tool relative to gun trials for both attend ($M_s = -1.77$ and $-1.41$, respectively; $t(2406) = -2.09, p = .037$) and suppress ($M_s = -2.22$ and $-1.19$, respectively; $t(2406) = -4.45, p < .001$), whereas there is no difference between targets for reappraise ($M_s = -2.95$ and $-3.01$, respectively; Figure 2d).

These results qualified a significant three-way interaction of Instructions Block x Prime x Target ($F(2, 2406) = 14.68, p < .001$). In an effort to unpack this complicated three-way interaction, we first examined the Prime x Target interaction separately within each level of block instructions (with the same random coefficients specified previously). The interaction of prime with target was significant for both attend ($F(1, 1113) = 9.26, p = .002$) and suppress ($F(1, 639) = 21.53, p < .001$), but not reappraise ($F < 1, p > .6$). As reported in Study 1, the Prime x Target interaction during attend reflects the large discrepancy between gun and tool trials for black, but not white, primes (see Figure 2d). Relative to all other trials types, the ERN amplitude was enhanced during black-tool trials, and subdued during black-gun trials (Figure 2d), which is consistent with expectations regarding the WIT. For those who performed expressive suppression while completing the WIT, the Prime x Target interaction was indicative of the rather diminished ERN observed during white-gun trials compared to both white-tool ($t(654) =$
-6.66, p < .001) and black-gun \((t(654) = -6.23, p < .001; \text{Figure 2d})\). Except for white-gun, no differences were found between the other WIT conditions and one another \((F_s < 1, ps > .7; \text{Figure 2d})\). However, for participants who performed reappraisal, there was only a main effect of prime \((F(1, 654) = 32.3, p < .001)\), suggesting that ERNs were greater when the prime was black relative to white, irrespective of target \(\text{(Figure 2d)}\).

To investigate specifically how block instructions interacted with race within this three-way interaction, a MLM analysis of Block Instructions x Prime was conducted separately within each Target type. A significant Instructions x Prime interaction was found across both tool \((F(2, 769) = 8.10, p < .001)\) and gun \((F(2, 787) = 22.97, p < .001)\) trials. Simple effects analyses indicate that reappraise elicited greater ERNs during black-gun trials relative to attend \((t(787) = 2.07, p = .039; \text{Figure 2d})\); this same effect was found to a marginal extent for black-tool trials \((t(769) = 1.86, p = 0.063)\). These effects are present as the result of the enhanced ERNs for the reappraise group during black-gun and -tool trials \(\text{(Figure 2d)}\). Additionally, the suppress group exhibited lower ERNs during white-gun trials compared to reappraise \((t(787) = -2.13, p = .033; \text{Figure 2d})\), which is also contributing to this three-way interaction.

**Study 2 Discussion**

The purpose of Study 2 was to address a major limitation in the design of Study 1, namely, that participants always completed the attend block first (in order to eliminate carryover effects that might occur if emotion regulation blocks were completed first), followed by reappraise and suppress, which were counterbalanced as the second and third blocks. This was necessary in order to rule out alternative hypotheses that may have contributed to the results obtained in Study 1, such as order effects and fatigue. This
problem was dealt with in Study 2 by facilitating comparisons of different block instructions (attend, reappraise, and suppress) between subjects instead of within. The same hypotheses tested in the within-subjects design in Study 1 were tested again in this between-subjects design.

First, we hypothesized that successful emotion regulation of error-related (aversive) negative affect would reduce motivation to change behavior on future trials, and result in participants making more errors overall. Contrary to expectations, block instructions did not differ in accuracy rates. Second, we hypothesized that reappraising negative emotions caused by making errors would increase error rates specifically in a way that is indicative of an increase in the behavioral expression of racial bias during the WIT. Although the typical WIT Prime x Target interaction did emerge in the accuracy data, it did not change across block instruction groups, as evidenced by the non-significant three-way interaction of prime, target, and block instructions. This finding also indicates that there is no difference between blocks in the black- and white-tool discrepancy. Additionally, automatic processes were not found to differ as a function of block instructions. Taken together, these results provide no evidence that cognitively reappraising or expressively suppressing emotional responses to errors made during the WIT decreases accuracy performance or increases racially biased responding.

Our third hypothesis predicted that an overall decrease in regulatory control, as measured by RT and PDP (i.e., control processes). In the present between-subjects analysis, block instructions had no effect on controlled processes. However, a marginal three-way interaction of block instructions with prime and target was found in the RT data, driven by differences in the Prime x Target interaction across block instructions.
The same relative pattern of WIT conditions generating this observed two-way interaction appears consistently across the blocks and is standard for the WIT, suggesting that participants required more time to ensure a correct response when a black prime preceded a tool compared to when it preceded a gun. This is consistent with the assumption that making correct responses during black-tool trials require overriding an automatic response. Although the same two-way interaction pattern is present during each of the block instructions, the effect size is (marginally, as implicated by the marginality of the three-way interaction) more robust in suppress relative to both reappraise and attend. This is actually the opposite of what we might expect if participants were lacking motivation to respond correctly (especially during black-tool trials), as we had predicted. Instead, this might mean that participants were taking more time to respond in a way that was less biased (albeit, this doesn’t translate into better performance, according to the accuracy data). Alternatively, the added task of having to inhibit expressions of emotion while completing the WIT might have exaggerated the time necessary to make correct responses, especially during trials that required overriding an automatic (biased) response. This seems like the more likely explanation of the present results, given that it fits better with the accuracy data since this alternative does not imply participants would be perform better.

Returning to the overall goal of these analyses, these results converge to provide no evidence that suggests reappraising errors produces any changes in the execution of cognitive control. Although the evidence regarding suppress appears to be mixed, the RT and PDP findings are not necessarily in conflict with one another. RTs can be altered for many reasons other taking time to respond more carefully. As such, the results from the
PDP analysis of control processes ought to carry more weight in providing evidence for this hypothesis, which does not support any changes in control implementation by applying expressive suppression to error-induced affect.

Lastly, our fourth hypothesis proposed that the reduction in distress from error commission by application of a successful regulatory strategy would be evident in the decreased amplitude of the ERN. Because reappraise has been shown to be a more effective emotion regulation strategy (e.g., Gross & Levenson, 1993), we thought that the ERN amplitude would be attenuated more during reappraise than suppress. As outlined previously in the discussion of Study 1 results, we differentiated between distress caused by the aversion of making errors more generally, and distress related specifically to feelings of compunction that arise after making errors indicative of racial bias.

Contrary to expectations, we found no main effect of block instructions, implying that there was no overall reduction in aversive feelings following the commission of errors. What differences we did find between blocks were specific to certain WIT conditions, as each block instructions produced differential patterns of ERN responses to the WIT conditions. As described previously under Study 1, the pattern of responses during the attend block was consistent with expectations regarding the WIT. That is, there was a large discrepancy between gun and tool trials for black, but not white, primes. Relative to white prime trial types, the ERN amplitude was enhanced during black-tool trials, and subdued during black-gun trials. However, the nature of this Prime x Target interaction apparent in attend was substantially altered by both emotion regulation strategies in Study 2, possibly indicating variation in the experience of compunction across blocks.
For the suppress subjects, the ERN during white-gun trials was diminished, while all other WIT trial types (including black-tool) were no different from one another. This could mean that participants did not feel any more distress after making a bias-related error (i.e., during black-tool trials) than they experienced after making a different type of error (e.g., during white-tool trials). This suggests that those participants may have been experiencing less bias-specific compunction compared to those who completed the attend block.

On the other hand, for participants in Study 2 who performed reappraisal, ERN amplitudes were greater when the prime was black relative to white, irrespective of target. Remarkably, these enhanced black prime ERN amplitudes were even larger than those produced by the attend block participants (Study 1). Given that the ERN during black-tool trials in particular was more negative for participants who reappraised errors relative to attend, one could argue that these participants may have experienced more compunction than the attend participants. However, this may not necessarily be the best interpretation, since it was not isolated to black-tool trials. That is, the ERN during reappraise was also enhanced during black-gun trials as well black-tool trials, and so it seems that this may not reflect compunction exactly.

On the whole, the results from the between-subjects approach to investigating the differential impact of regulating emotional response to errors on the ERN, performance, and indices of control do not converge. The ERN results provide some evidence to suggest that at best, suppress may mitigate the compunction from making biased errors. Meanwhile, reappraise appears to have actually increased distress during specific trial types (black primes), and also compunction as well. Both of these findings speak against
our a priori hypotheses (possible explanations are considered in depth in the General Discussion section below). If our a priori assumption on which these interpretations of the ERN data are based is correct, and these fluctuations in ERN do in fact reflect changes in affect, then we might also expect to observe related behavioral changes evident in performance during the WIT. However, the only difference in behavior observed between blocks was RT data, which showed a marginally more biased pattern of responding. This pattern did not transfer to other measures of performance, and so it is difficult to derive much meaning from it. All other indices of performance and control converge, in the sense that they all provide no evidence that either suppress or reappraise had any impact.

**General Discussion**

Although extensive research has been done on the separate topics of emotion regulation and controlling the expression of implicit racial bias, little work has been done to characterize overlap between these two self-regulatory processes. In an attempt to bridge these two areas of research, the purpose of the current experiments was to investigate the impact of emotion regulation strategies on the inhibition and expression of implicit racial bias. Specifically, we sought to test the effect of regulating the distress following the accidental expression of automatic bias on behavioral and neural manifestations (e.g., the ERN) of implicit racial attitudes. In two EEG studies, expressive suppression and cognitive reappraisal were applied to reduce the distress experienced following erroneous responses during the WIT (Payne, 2001) – a task designed to measure implicit racial bias. These two strategies were compared, alongside a control condition, in both a within- (Study 1) and between-subjects design (Study 2). Although
both studies sought to answer the same questions, and employed very similar methods, the results obtained in both studies were very different from one another.

As already described in other sections of this paper, we first sought to demonstrate that regulation of aversive feelings produced by error commission would reduce motivation to change behavior on future trials, and reduce performance in general. The second goal was to show that feelings of compunction that arise only after committing errors indicative of the accidental expression of racial bias could also be reduced by means of emotion regulation. Doing so was expected to impact race-bias specific neural and behavioral outcomes, such that the behavioral expression of racial bias would increase as the result of decreased race-bias specific ERN, which is thought to have implications for maintaining control. We predicted that cognitive reappraisal would be successful at regulating both the aversive feeling caused by errors in general, as well as when the source of error-related distress comes specifically from compunction. For expressive suppression, previous research mostly suggests that this strategy would most likely have little or no effect on behavioral and emotional outcomes, although there was also a possibility that this strategy could be successful (see Goldin et al., 2008).

Consistent with our hypotheses, the behavioral results from Study 1 demonstrate that reappraising errors during the WIT effectively reduced cognitive control engagement, thus leading to decreased accuracy overall and increased expression of racial bias. Although the increase in racial bias would have been expected to result from reduced compunction, the ERN data repudiate this claim. The ERN analyses indicate that reappraise was only effective at reducing the distress caused by making errors in general. Like reappraise, suppress also caused a decrease in cognitive control, which was also
evident in the reduced performance accuracy. Although these behavioral outcomes were actually more pronounced in suppress than reappraise, the ERN results suggest that reappraise was more effective than suppress at reducing general error-related distress. Additionally, like reappraise, the ERN data also show that suppress appears to be ineffective when this distress is coupled with the feeling of compunction following a biased error. Although the behavioral and neural results don’t entirely converge in terms of identifying which regulation strategy is superior within the context of attenuating emotional response to errors, it seems clear that both strategies were at the very least effective at reducing the distress related to errors.

While most aspects of the behavioral and neural data from the within-subjects analysis supported our main hypotheses (namely, those regarding the efficacy of reappraisal when applied within this paradigm), the vastly divergent results obtained from the between-subjects design suggest that ultimately the differences observed in Study 1 may be attributable to the order in which the blocks occurred. It is suspected that this may be related to fatigue. Regardless of the reason, it seems that completing reappraise or suppress during the second or third block has differential (and in some cases, opposite) effects compared to performing these tasks during the first block. Since the between-subjects comparison resolves the order problem from Study 1, these results outweigh those from Study 1.

The between-subjects ERN results provide evidence suggesting that suppress may mitigate compunction elicited by making biased errors (without reducing the overall distress from making errors in general. This is the complete reversal of the effect of suppress on the ERN found in Study 1 (i.e., decreased overall distress and no reduction in
compunction). Meanwhile, reappraise appears to have actually increased distress (measured by the ERN) during specific trial types (black primes), which means an increase in compunction as well. This effect is in the opposite direction as the ERN results obtained in the within-subjects design. These findings are counter to our hypotheses predicting that reappraise would reduce the ERN, and suppress would most likely be ineffectual. On the other hand, most of the behavioral indices of performance, control, and bias converge with one another, providing a lack of evidence that either suppress or reappraise impacted behavioral outcomes. This also contradicts our hypotheses, and also doesn’t seem to fit with the group differences observed in the neural measures.

This discrepancy in the ERP data and behavioral data may seem inconsistent, especially for a neural component such as the ERN, which has been previously linked to measures of performance and control (e.g., Bartholow et al., 2012). One possible reason for this seeming inconsistency might be that the measures of behavior used here are less sensitive than ERP measures, and thereby failed to capture any differences that might be evident. Additionally, it is possible that changes in emotional states may have occurred without culminating in any observable behavioral effects.

The present behavioral data from the between-subjects design seem to point to the more obvious explanation that these particular types of emotion regulation strategies simply have no effect on accuracy and control. However, this explanation contradicts the extant literature on reappraisal, most of which claim that it is effective when applied in a variety of contexts, including error processing (Hobson, Saunders, Al-Khindi, & Inzlicht, 2014; Ichikawa, Siegle, Jones, Kamishima, Thompson, et al., 2011).
Alternatively, it is more likely that the effect on behavior is rather small (e.g., the sizes of effects on accuracy involving block instructions are estimated by $\eta^2_p$ to range from 0.01 to 0.02), and therefore, the present experiment would not have had enough power to detect these small differences. A small effect size is not surprising, given the low level of emotional arousal associated with feelings of distress caused by making errors during a cognitive task in the lab. This is especially true when the affective response induced during this task is compared to higher arousal levels induced by stimuli (e.g., graphic emotional pictures) typically used in emotion regulation tasks. If the effect really does exist, then effect sizes that are as small as these would require a much larger sample size than a minimum of 30 participants per group in order to properly test this effect using a between-subjects design. Some small effects can have important implications, especially when they affect entire populations (e.g., implicit racial bias; Greenwald, Banaji, & Nosek, 2014). However, realistically speaking, this particular small (or non-existent) behavioral effect would most likely not have significant implications that would affect an entire population.

Another possibility that might explain these unexpected null results is that this particular brand of reappraisal (i.e., cognitive reappraisal) may have been an ill-suited reappraisal technique for this particular paradigm. For example, participants may have not had enough time to fully and effectively apply this strategy following each and every error due to the very brief inter-trial-interval (average duration of 1 sec). While participants acknowledged that this made the task more challenging, many of them nonetheless claimed they were still able to implement the reappraisal strategy following errors most of the time (according to responses during post-experiment questionnaires...
and debriefing). Although self-report and self-insight are not always reliable (Nisbet & Wilson, 1977), it still seems more plausible that the sample size is too small to adequately test between-group differences.

Although these results were not what we were expecting, they do provide an opportunity to refute the alternative hypothesis regarding cognitive load mentioned earlier. Since performance did not change as a function of block instructions, this would imply that the combination of performing emotion regulation strategies while simultaneously completing the WIT is not taxing cognitive resources.

Returning to the between-subjects ERP results, we found that suppress altered the ERN response pattern across WIT conditions, such that these participants appeared to be experiencing less compunction in response to biased errors. This is not the first time that suppress has been found to be effective at reducing affect (e.g., Goldin et al., 2008). However, it is strange that a reduction in compunction would not be accompanied by reduction in aversion to errors more generally. Regardless, there is still the possibility that in spite of all the evidence claiming that suppress can be ineffective and/or counterproductive (e.g., Gross & John, 2003; Gross, 2003), perhaps it is necessary to consider the possibility that suppress might sometimes be better suited as an emotion regulation technique under certain circumstances.

Alternatively, another explanation is that expressive suppression provided a distraction task that was sufficiently challenging enough that it required participants to shift their attention away from the emotional consequences of making racially biased errors (Goldin et al., 2008). Although they may have been distracted from one aspect of the erroneous response, they were required by the suppression task itself to attend to
errors. This could explain why we observed a reduction in one source of error-induced distress (i.e., compunction), without observing a reduction in the aversive affect produced by making errors in general.

Additionally, we found that reappraisal actually increased the amplitude of the ERN during black prime trials, which is once again, inconsistent with our predictions. One explanation is that the particular reappraisal strategy (i.e., cognitive reappraisal) selected for this experiment may have caused participants to actually be more focused on errors, which may have increased their emotional response. If this is true, then applying a different reappraisal strategy (e.g., situational reappraisal, which involves taking a detached perspective; e.g., Gross, 1998) to this paradigm might have yielded results that were more consistent with our hypotheses. However, if participants were hyper-focused on errors, then we might also expect this to cause changes in behavioral measures of performance. Additionally, it is unclear why this would have generated increased ERN amplitudes during only two out of the four WIT condition trial types. As such, this alternative hypothesis does not seem to fully explain the present data.

Another alternative explanation for this unexpected finding draws from recent fMRI research demonstrating that cognitive reappraisal actually increases activation in the dACC (for meta-analysis, see Buhle, Silvers, et al., 2013). It is thought that the conflict monitoring system of the dACC is recruited to track one’s emotional state while reappraising in order to gauge whether more or less regulation is necessary (Ochsner, Silvers, & Buhle, 2012). Additional work has shown that the anterior cingulate structures recruited by reappraisal activation overlap with those recruited during emotional conflict detection (Etkin, Egner & Kalisch, 2011), like that which occurs during stereotype-
incongruent trials in the WIT. The increased ERN amplitude observed during reappraisal may have resulted from increased activity in the dACC. Here, it seems that the dACC conflict monitoring/distress system may have been recruited to track emotional states during reappraisal, resulting in larger ERN amplitudes. This could explain why the ERNs were especially enhanced during black prime trials in particular, where more monitoring may have been necessary.

**Limitations**

Naturally, the present experiments presented here are not without limitations, some of which have already been mentioned. Among the limitations not discussed thus far is the design of the block 1 analysis, in that comparisons were made across two experiments. Technically, participants were not randomly assigned to regulation groups in the between-subjects comparison, because all participants in Study 1 were assigned to complete attend during their first block. This may have allowed the influence of unanticipated variables confounded with group assignment (e.g., time of participation). Additionally, there are many differences in the designs of studies 1 and 2 that may have contributed to between-group differences. For example, the attend group completed three blocks, whereas Suppress and Reappraise groups (Study 2) only completed two blocks. There is evidence suggesting that the knowledge of having more or fewer blocks remaining may have affected performance during the first block (Muraven, Shmueli, & Burkley, 2006). Although Study 2 was designed to address the order effect of block instructions, because of these discrepancies in the procedures for Studies 1 and 2, Study 2 may not have fully addressed this problem.

**Implications and Future Directions**
To our knowledge, reappraisal has only been applied to error processing in two other studies to date (Hobson et al., 2014; Ichikawa et al., 2011), and in the current experiments, this was extended to errors indicative of racial bias. Additionally, the present experiments represent the first attempt to apply expressive suppression to cope with error-related affect. More research is necessary to clarify how reappraisal and other emotion regulation strategies affect neural and behavioral aspects of error monitoring. Additionally, given how crucial the block order effect was in determining the neural and behavioral outcomes, this has important implications for emotion regulation research using within-subjects designs. However, because this was not a primary aim of the present study, we were unable to fully investigate the interaction of block order with emotion regulation strategy instructions with a sufficiently powered sample (see supplementary results for an exploratory analysis of this interaction). Future research should seek to further investigate how fatigue and/or block order impact performance during emotion regulation tasks.
Table 1

*Within-Subjects (Study 1): Mean Accuracy Rates and Response Times in the Weapons Identification Task (WIT)*

<table>
<thead>
<tr>
<th>Instructions block and prime type</th>
<th>Accuracy Rates</th>
<th>Response Times</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Target type</td>
<td>Target type</td>
</tr>
<tr>
<td></td>
<td>Gun</td>
<td>Tool</td>
</tr>
<tr>
<td></td>
<td>Gun</td>
<td>Tool</td>
</tr>
<tr>
<td><strong>Attend block</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black primes</td>
<td>.78 (.14)</td>
<td>351.09 (51.75)</td>
</tr>
<tr>
<td>White primes</td>
<td>.70 (.15)</td>
<td>359.88 (54.24)</td>
</tr>
<tr>
<td><strong>Reappraise block</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black primes</td>
<td>.76 (.14)</td>
<td>323.43 (69.73)</td>
</tr>
<tr>
<td>White primes</td>
<td>.61 (.16)</td>
<td>333.39 (76.2)</td>
</tr>
<tr>
<td><strong>Suppress block</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black primes</td>
<td>.73 (.16)</td>
<td>308.21 (70.61)</td>
</tr>
<tr>
<td>White primes</td>
<td>.60 (.15)</td>
<td>316.63 (74.1)</td>
</tr>
</tbody>
</table>

*Note:* Values in parentheses are standard deviations.
Table 2

*Between-Subjects (Studies 1 and 2): Mean Accuracy Rates and Response Times in the WIT*

<table>
<thead>
<tr>
<th>Instructions block and prime type</th>
<th>Accuracy Rates</th>
<th>Response Times</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Target type</td>
<td>Target type</td>
</tr>
<tr>
<td></td>
<td>Gun</td>
<td>Tool</td>
</tr>
<tr>
<td></td>
<td>Gun</td>
<td>Tool</td>
</tr>
<tr>
<td><strong>Attend block (Study 1)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black primes</td>
<td>.78 (.14)</td>
<td>351.09 (51.75)</td>
</tr>
<tr>
<td>White primes</td>
<td>.70 (.15)</td>
<td>359.88 (54.24)</td>
</tr>
<tr>
<td><strong>Reappraise block (Study 2)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black primes</td>
<td>.80 (.03)</td>
<td>366.66 (9.29)</td>
</tr>
<tr>
<td>White primes</td>
<td>.73 (.03)</td>
<td>372.37 (1.75)</td>
</tr>
<tr>
<td><strong>Suppress block (Study 2)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black primes</td>
<td>.83 (.02)</td>
<td>363.22 (8.62)</td>
</tr>
<tr>
<td>White primes</td>
<td>.71 (.03)</td>
<td>378.05 (8.84)</td>
</tr>
</tbody>
</table>

*Note:* Values in parentheses are standard deviations.
Figure 1. Study 1 performance across block instructions (within subjects). (a) Displayed here are the mean accuracy rates for Block Instructions x Prime x Target (displayed on the x-axis). The y-axis represents the mean proportion of accurate trials out of the total number of trials where responses were given (i.e., excluding missed trials). (b) Automatic and controlled processes influencing responses were discerned via PDP. The y-axis represents the PDP scores, which are intended to reflect the mean level of controlled or automatic process that contributed to the observed accuracy behavior. PDP = process dissociation procedure; PDCtrl/Ctrl = controlled processes; PDAuto/Auto = automatic processes. Error bars represent the standard error of the mean.
Figure 2. Within- and between-subjects comparison of the effect of emotion regulation strategies on the ERN. (a) Montage of all the electrode locations from which recordings were taken during Studies 1 and 2. Locations of channels included in analysis of ERN
(F1, F2, Fz, FC1, FC2, FCz, C1, C2, and Cz) are highlighted in yellow. (b) Analysis of the ERN revealed an interaction of Instructions x Prime. Error bars represent the standard error of the mean. Note that the y-axis scale has been reversed (i.e., negative is up and positive is down) to be consistent with the standard method of displaying ERPs. (c) Composite of the nine electrodes highlighted in (a) during error trials for Study 1 (within-subjects comparison of block instructions). Grand averages were weighted by the number of trials and subjects per condition, so as to better represent the MLM estimated marginal means. Highlighted in purple is the time interval during which the ERN averages were quantified (30 – 110 ms). Error responses occurred at 0 ms. (d) Composite of the weighted grand averages at the same nine electrodes during the first block of Studies 1 and 2 (between-subjects comparison of block instructions). ERN = error-related negativity. ERP = event-related potential.
Figure 3. Between-subjects comparison of block instructions on accuracy rates (a) and automatic and controlled processes (b). Displayed are the means from the first block completed by participants in Studies 1 and 2 (i.e., attend participants were in Study 1, and suppress and reappraise participants were in Study 2). Automatic and controlled processes were derived via PDP. PDP = process dissociation procedure.
References


Supplementary Material

Block Order Effect (Studies 1 and 2): Results and Discussion

In order to probe the effect of order of block instructions, the first and second blocks were compared across Studies 1 and 2. Participants from Study 1 were split into two groups based on which emotion regulation strategy they completed during the second block. Those for whom attend was the first block and suppress was the second block in Study 1 (which we will now call group 1; \( n = 24 \)) were compared with participants in the suppress group in Study 2 (i.e., block 1 was suppress and block 2 was attend; we will now refer to them as group 2; \( n = 30 \)). This created a 2 (Block Order Group: group 1, group 2) \( \times \) 2 (Block Instructions: attend, suppress) fully crossed, mixed factorial design, with order group as a between-subjects factor. A similar design was generated for reappraise: participants in Study 1 who performed reappraise during the second block (group 3; \( n = 22 \)) were compared to those in the reappraise group in Study 2 (group 4; \( n = 33 \)). Here, we are assuming that the knowledge of having to complete a third block did not affect performance during the first or second block for participants in Study 1.

**Accuracy.** Data from the suppress groups (1 and 2) were analyzed in a 2 (Order Group: group 1, group 2) \( \times \) 2 (Block Instructions: attend, suppress) \( \times \) 2 (Prime) \( \times \) 2 (Target) fully crossed, mixed factorial ANOVA. A significant two-way interaction of Prime \( \times \) Target \( (F(1, 52) = 70.42, p < .001, \eta^2_p = 0.58) \) reflects the typical WIT pattern discussed previously. Additionally, the results showed a significant interaction for Block Instructions \( \times \) Order Group \( (F(1, 52) = 30.51, p < .001, \eta^2_p = 0.37) \). During the first block, participants were more accurate during suppress than attend, whereas when suppress was
the second block (i.e., following attend), the effect was reversed and accuracy decreased. Performance during attend was unaffected by order (Figure S2a).

The data from the reappraise groups were analyzed in a separate 2 (Order Group: group 3, group 4) x 2 (Block Instructions; attend, reappraise) x 2 (Prime) x 2 (Target) full factorial ANOVA, which showed previously discussed, typical effect for target \(F(1, 53) = 6.76, p = .012, \eta_p^2 = 0.11\) and Prime x Target interaction \(F(1, 53) = 57.73, p < .001, \eta_p^2 = 0.52\). A marginal two-way interaction for Instructions x Prime was found \(F(1, 53) = 3.75, p = .058, \eta_p^2 = 0.07\) suggesting that during attend, accuracy was better for black compared to white \((\Delta = .02)\) primes, whereas there was no difference during reappraise \((\Delta = 0.0);\) Figure S2c). Similar to suppress, here we also found an effect of Instructions x Order Group \(F(1, 53) = 5.14, p = .028, \eta_p^2 = 0.09\), although the effect size is notably smaller for reappraise. Like suppress, overall accuracy was improved when reappraise was the first block, and reduced when it was the second block, whereas no such block order difference is observed in attend (Figure S2c). Moreover, a four-way interaction of Order Group x Instructions x Prime x Target emerged \(F(1, 53) = 9.22, p = 0.004, \eta_p^2 = 0.15\), which appears to be mostly driven by decreased accuracy during black-tool and white-gun trials when reappraise was the second block (see Figure S2c).

**RT.** RT data from the suppress groups was analyzed using a 2 (Order Group: group 1, group 2) x 2 (Block Instructions: attend, suppress) x 2 (Prime) x 2 (Process) fully crossed, mixed factorial ANOVA. This revealed significant effects for target \(F(1, 52) = 12.41, p = 0.001, \eta_p^2 = 0.19\) and Prime x Target \(F(1, 52) = 47.16, p < 0.001, \eta_p^2 = 0.48\), similar to task-typical patterns discussed earlier. Additionally, there was a significant interaction of Instructions x Order Group \(F(1, 52) = 34.47, p = 0.004, \eta_p^2 = \)
There appears to be an order effect (across groups), such that RTs during the second block were faster (averaged across groups, $\Delta = 33.0$ ms). Whether attend occurred during the first or second block made no difference in terms of reaction time ($\Delta = 14.69$). However, when participants performed suppress during the first block, they were much slower to respond compared to those who suppressed during the second block ($\Delta = 51.30$).

For the reappraise groups, a similar $2 \times 2 \times 2 \times 2$ mixed factorial ANOVA exposed significant effects for target ($F(1, 53) = 21.09, p < 0.001, \eta^2_p = 0.29$), prime ($F(1, 53) = 6.09, p = 0.017, \eta^2_p = 0.10$), and Prime x Target ($F(1, 53) = 50.81, p < 0.001, \eta^2_p = 0.49$), which are in line with effects described earlier. Like suppress groups, there was also a significant interaction of Instructions x Order Group ($F(1, 53) = 21.80, p < 0.001, \eta^2_p = 0.29$), indicating that subjects were faster at responding during the second block across groups, especially the subjects for whom reappraise was the second block (group 3: attend $M = 368.84$, reappraise $M = 346.13$; group 4: attend $M = 356.57$, reappraise $M = 373.59$). There is also a marginal three-way interaction of Instructions x Prime x Target ($F(1, 53) = 3.83, p = 0.056, \eta^2_p = 0.07$), which suggests that the Prime x Target interaction discussed previously is more pronounced for attend compared to reappraise.

**PDP.** For the suppress groups, a $2$ (Order Group: groups 1 and 2) x $2$ (Block Instructions: attend, suppress) x $2$ (Prime) x $2$ (Process) mixed factorial ANOVA revealed main effects for prime ($F(1, 52) = 47.73, p < .001, \eta^2_p = 0.48$), process ($F(1, 52) = 10.39, p = .002, \eta^2_p = 0.17$), and an interaction for Prime x Process ($F(1, 52) = 60.86, p < .001, \eta^2_p = 0.54$) that were consistent with findings discussed earlier. A Block
Instructions x Order Group interaction \((F(1, 52) = 30.43, p < .001, \eta_p^2 = 0.37)\) appears to be caused primarily by variability in control processes (since automatic appears to remain constant throughout; see Figure S2b). As such, it simply echoes what is captured in the three-way Block Instructions x Order Group x Process interaction \((F(1, 52) = 13.11, p = .001, \eta_p^2 = 0.20)\), signifying that participants exhibited more control during suppress when it was the first block compared to attend, and less control when suppress occurred as the second block (Figure S2b), whereas attend does not differ as a function of block order. Additionally, there was no variation in automatic processes across blocks.

An analogous 2 x 2 x 2 x 2 ANOVA conducted with the reappraise groups showed an effect for prime \((F(1, 53) = 60.77, p < .001, \eta_p^2 = 0.53)\), a marginal effect for process \((F(1, 53) = 3.90, p = .054, \eta_p^2 = 0.07)\), and a significant interaction of Prime x Process \((F(1, 53) = 48.87, p < .001, \eta_p^2 = 0.48)\). Akin to the results above, there were interactions for Block Instructions x Order Group \((F(1, 53) = 15.23, p < .001, \eta_p^2 = 0.22)\), which demonstrates an overall decrease across both automatic and controlled processes while performing reappraise when it was the second block, whereas these processes during attend were unaffected by block order (Figure S2d). A significant three-way interaction of Block Instructions x Order Group x Prime \((F(1, 53) = 11.08, p = 0.002, \eta_p^2 = 0.17)\) suggests that this reduction in PDP scores during reappraise as the second block primarily occurs for the white prime, given that there is little fluctuation across blocks for the black prime (see Figure S2d).

**ERN.** The ERN amplitude were quantified according to the previously mentioned procedures for the 3 x 3 frontal-midline electrodes centered on FCz. Suppress and reappraise groups were analyzed separately in 2 (Block Instructions) x 2 (Block Order: 1st
block, 2\textsuperscript{nd} block) x 2 (Prime) x 2 (Target) MLMs with random intercepts for subject and electrode nested within subject.

The analysis conducted with the suppress group data yielded main effects for target ($F(1, 2654) = 75.29, p < .001$) and prime ($F(1, 2654) = 14.87, p < .001$), block order ($F(1, 2654) = 4.45, p = .035$), and instructions ($F(1, 2654) = 11.86, p = .001$), indicating ERN amplitude was greater during the first block relative to the second block, and attend relative to suppress (see Figure S3a). A two-way interaction of Instructions x Target ($F(1, 2654) = 9.03, p = .003$) suggested that gun trials elicited similarly sized ERNs during both attend ($M = -0.99$) and suppress ($M = -0.93$), but on tool trials, ERNs during attend ($M = -2.38$) are greater than suppress ($M = -1.60$). There was also an interaction of Prime x Block Order ($F(1, 2654) = 3.91, p = .048$), such that ERN amplitude was greatest during black trials while completing the first block ($M = -1.95$), which subsequently decreased while completing the second block ($M = -1.46$), whereas there block order did not seem to affect ERNs during white trials ($M = -1.26$ during block 1, $M = -1.24$ during block 2). There was a Target x Instructions x Block Order interaction ($F(1, 2654) = 8.55, p = .004$), which is caused by the convergence of gun and tool ERNs when suppress was performed during the second block ($\Delta = 0.24$), which is distinct from the large gun-tool difference present in all other blocks ($d$s range from 1.12 to 1.65). There was a Prime x Target x Instructions ($F(1, 2654) = 6.15, p = .013$), suggesting that attend produces larger ERNs across all WIT conditions, except for black-gun, where suppress appears to be larger (Figure 9). Additionally, we found a Prime x Target x Block Order interaction ($F(1, 2654) = 10.32, p = .001$). Although overall ERN amplitude decreases from block 1 to block 2, this does not appear to occur for black-tool
nor white-gun trials. (In other words, the black-tool and white-gun conditions are immune to the diminishing effects of block order.) As such, the difference between black-tool and both white-tool and black-gun is greater during the second block, while the difference between gun conditions disappears.

Finally, the present analysis identified a significant four-way interaction of Block Order x Instructions x Prime x Target ($F(1, 2654) = 15.69, p < .001$). The ERN during white-gun trials appears to be relatively stable across all blocks (Figure S3a). The same is true for black-gun ERNs, except when participants were performing suppress during the first block, which seemed to elicit larger ERNs in that condition. For both black- and white-tool, there is little variation in the ERN signal across blocks, except for the drastic decrease that occurs when suppress occurs as the second block. It seems that the source of this four-way interaction are the changes that occur when suppress is the first versus second block, since attend is relatively unaffected by block order (Figure S3a).

When participants engaged in expressive suppression during the first block, there is a small, but notable decrease in both black-tool and white-gun ERN amplitude, and an increase in black-gun amplitude relative to attend. Unlike attend, suppress generates no differences between black-tool and white-tool or black-gun. Instead, the relatively smaller ERN amplitude during white-gun creates the differences observed within this block (Figure S3a). When suppress was completed during the second block, ERN amplitudes were much smaller during both tool conditions. However, similar to attend during the first block, the ERNs were largest during black-tool, thus creating a sizeable gap between both black-gun and white-tool (Figure S3a). In summary, it seems that this four-way interaction is primarily due to differences in the WIT condition interactions.
observed within suppress as the first block, and the overall decrease in ERN amplitude during tool trials when it suppress was the second block.

The 2 (Block Instructions) x 2 (Block Order: 1st, 2nd block) x 2 (Prime) x 2 (Target) MLM performed on data from the reappraise groups revealed main effects for both prime \( (F(1, 2821) = 9.40, p = .002) \) and block instructions \( (F(1, 2821) = 4.25, p = .040) \), meaning that overall ERN amplitude was greater during black compared to white trials, and also attend blocks relative to reappraise (Figure S3b). A marginally insignificant effect of block order \( (F(1, 2821) = 3.50, p = .062) \) suggests that there was a trend in the direction of block 1 eliciting slightly more negative ERNs than block 2 (Figure S3b). A two-way interaction of Prime x Target \( (F(1, 2821) = 27.75, p < .001) \) reflects that black-tool ERN amplitude is larger than both white-tool and black-gun, as mentioned in previous analyses (Figure S3b); there also appears to be a difference between white-tool and white-gun (Figure S3b). There was a significant two-way interaction of Prime x Instructions \( (F(1, 2821) = 53.83, p < .001) \). Reappraise reflects the main effect of prime, where black primes produce larger ERNs than white primes – an effect typical of the WIT. However, during attend, this pattern is reversed, such that white primes \( (M = -2.54) \) are associated with larger ERNs than black primes \( (M = -2.07) \), which is somewhat unusual (Figure S3b). Another two-way interaction of Block Instructions x Block Order \( (F(1, 2821) = 4.87, p = .027) \) suggests that ERN amplitudes were generally lower during reappraise when it was the second block \( (M = -1.07) \) compared to the first \( (M = -3.09) \), as might be expected. However, this pattern is reversed for attend, so that ERNs are actually smaller when attend was the first \( (M = -1.51) \) compared to the second block \( (M = -3.11) \).
Like suppress, there was a three-way interaction of Prime x Target x Block Order ($F(1, 2821) = 8.56, p = .004$), although it represents a different pattern in the present data. During the first block, black-tool, black-gun, and white-gun are equivalent ($M$s = -2.42, -2.46, and -2.44, respectively), while white-tool ($M = -1.88$) is smaller, generating a small difference between white-tool with both white-gun and black-tool. However, during the second block, the ERNs decrease for both black-gun ($M = -1.79$) and white-tool ($M = -1.49$), white-gun stays about the same ($M = -2.30$), while the ERNs during black-tool ($M = -2.79$) actually increase on average. As such, the between-condition differences observed during block 1 are more pronounced during block 2, and a difference between black-gun and both black-tool and white-gun emerges. The increased ERN for the black-tool condition observed during block 2 is consistent with normal WIT findings.

Lastly, there is a significant four-way interaction of Prime x Target x Instructions x Block Order ($F(1, 2821) = 12.80, p < .001$), which seems to be primarily driven by an effect of order group assignment. Participants who completed attend during the first block (i.e., group 3) exhibited much smaller ERNs overall compared to those who completed the instructions blocks in the reverse order (group 4). ERNs are much smaller when attend is the first block, where white-tool ($M = -2.34$) is greater than all other WIT conditions ($d$s = 1.10 to 1.13). When these same subjects subsequently performed reappraise during the second block, the ERN for black-tool actually increases ($M = -2.41$), while the ERNs for all WIT conditions is decrease (see Figure 9). However, when reappraise is the first block for subjects in group 4, they exhibited larger ERNs across all conditions, especially during black-tool and -gun trials ($M$s = -3.60 and -3.71, respectively). ERNs during white prime trials were smaller and did not differ from each
other ($M_s = -2.52$ and -.254, for tool and gun, respectively). When these participants later completed attend during the second block, their ERNs during both white-tool and -gun slightly increased ($M_s = -3.12$ and -3.48, respectively), while black-tool and -gun actually decreased ($M_s = -3.16$ and -2.69, respectively). The pattern that resulted during this attend block is actually similar to the interactions demonstrated while the other group of participants completed attend during the first block. Although the ERNs are much larger overall, again white-gun ($M = -3.48$) emerges as the largest ERN, and the divergence amongst the other three WIT conditions are diminished ($ds = 0.05$ to 0.48).

However, it is important to note that there is a sizeable difference between the two order groups, thus making it difficult to gleam meaning from the present results. Regardless, it does seem that ERNs associated with errors indicative of bias (black-tool) have larger amplitudes during reappraise compared to attend (across both groups). Similarly, the gap between black- and white-tool ERNs is also greater during reappraise as well, which is consistent with our predictions. However, both groups of participants exhibited atypical ERN patterns, such that white-gun elicited a larger ERN amplitude than other conditions. This makes the present results difficult to interpret.

**Error Positivity (P_E)**

The error positivity ($P_E$) is a positive-deflecting ERP component that occurs following erroneous responses, directly after the ERN (the peak is usually approximately 200 to 400 ms post-response). It is thought to originate from the more rostral region of the ACC (Herrmann et al., 2004). Although it often happens in conjunction with the ERN, it is thought to represent separate neural processes (Overbeek et al., 2005). Previously, the $P_E$ was posited to reflect one’s awareness of having made an error (e.g.,
Nieuwenhuis, Ridderinkhof, Blom, Band, & Kok, 2001). However, more recent evidence suggests it may reflect motivational salience of errors (; Davies et al., 2001; Ridderinkhof, Ramautar, & Wijnen, 2009; for review, see Overbeek et al., 2005). More research is still necessary to characterize the functional significance of the P_E.

Although the present experiments were not designed to manipulate P_E, initial inspection of the grand average waveforms for error trials (Figure S1) suggested that distinct patterns had emerged. However, hypotheses regarding the P_E were not generated a priori, and thus the results described here are exploratory. The primary purpose of conducting these analyses was to add to the somewhat limited body of literature on P_E (compared to the attention received by the ERN), with the hope that it may inform future research and theory regarding the significance of P_E. As with all exploratory analyses, these results should be interpreted with caution.

**Study 1: Results and discussion.** Like the ERN analyses, 2 x 2 condition cells per participant were excluded from analyses if they did not contain at least 5 (artifact-free) error trials. The P_E was quantified for each subject as the average signal between 150 to 250 ms following erroneous responses (same interval used by Bartholow et al. 2012). It was determined by visual inspection that Cz was the electrode site at which P_E peaked. As such, the analysis included the 3 x 3 swatch of central midline electrodes centered on Cz (FC1, FC2, FCz, C1, C2, Cz, CP1, CP2, and CPz).

The 3 (Block Instructions) x 2 (Prime) x 2 (Target) MLM analysis of P_E revealed main effects for prime ($F(1, 4165) = 15.10, p < .001$) and instructions ($F(1, 4165) = 38.57, p < .001$). Additionally, there were two-way interactions of Prime x Target ($F(1, 4165) = 11.06, p < .001$), and Instructions with both Target ($F(1, 4165) = 17.21, p < .001$)
and Prime \((F(1, 4165) = 10.95, p < .001)\). All of these effects were qualified by a significant three-way interaction of Instructions x Prime x Target \((F(1, 4165) = 3.48, p < .001)\).

To understand this complex interaction, we began by examining a 2 (Prime) x 2 (Target) MLM ANOVA within each block separately. The interaction of prime and target was significant during both attend \((F(1, 1113) = 27.88, p < .001)\) and reappraise \((F(1, 1185) = 11.67, p < .001)\), but did not quite reach significance during suppress \((F(1, 1068) = 2.99, p = .084)\). Examination of the waveforms (Figure S1a) for the attend block suggests that the P\(_E\) amplitudes for each condition are split by stereotype congruency. That is, the mean P\(_E\) signal for black-tool is comparable to white-gun (i.e., stereotype-incongruent conditions; Figure S1a), and the same is true for black-gun and white-tool (stereotype-congruent\(^8\)). During the attend block, the stereotype-incongruent trials (black-tool and white-gun) both elicited greater P\(_E\) than the stereotype-congruent trials (black-gun: \(t(1113) = 2.98, p = .003\) and \(t(1113) = 3.73, p < .001\), respectively; white-tool: \(t(1113) = 3.75, p < .001\), and \(t(1113) = 4.40, p < .001\); Figure S1a). Although reappraise also yielded a significant Prime x Target interaction, the simple effects pattern is different from attend (Figure S1a). Unlike attend, during reappraise, black-tool is not significantly different from black-gun or white-tool \((ps > .4)\). However, the divergence in primes when the target was a gun \((t(1185) = 4.10, p < .001)\), and in targets when the prime was a white

---

\(^8\) This is not to say that White people are stereotypically associated with tools. “Stereotype congruency” during the WIT is determined by the strength of the association of Black males and guns. Since Black males are more associated with guns than White males, by means of a contrast effect, this causes White males to become associated with tools within the context of this specific task (see Hilgard, Scherer, & Bartholow, In prep)
face ($t(1185) = 5.53, p < .001$) observed during attend was preserved during reappraise (Figure S1a).

Additionally, the differential effect of block instructions on $P_E$ in response to prime type was explored by conducting 3 (Instructions) x 2 (Prime) ANOVAs separately for each target. For tools, the results indicate that prime interacted with instructions ($F(2, 1741) = 10.17, p = .001$), whereas for guns, there is no such interaction ($F < 2, p > .2$). Simple effects analyses (for tools) revealed no differences in $P_E$ signal between blocks when primed with a Black face ($t < 1, p > .3$). However, when errors were made during tool trials after being primed with a White face, $P_E$ was more positive during reappraise compared to attend ($t(1741) = 2.81, p = .005$; Figure S1a), and even more so during suppress (attend: $t(1741) = 7.88, p < .001$; reappraise: $t(1741) = 5.36, p < .001$).

To summarize, it seems that this three-way interaction was produced by a combination of multiple differences across blocks. During attend, there was an effect for stereotype congruency, such that higher conflict trials (incongruent) generated a greater $P_E$ than lower conflict trials (congruent). However, this pattern is only partially evident during reappraise; like attend, $P_E$ during white-gun (incongruent) is greater than both congruent trial types, but not for the other incongruent condition (i.e., black-tool). During suppress, this interaction was not significant. Additionally, block differences were observed during white-tool trials, such that $P_E$ was more positive during reappraise than attend, and more so during suppress. Together, these factors contribute to the three-way interaction observed in $P_E$.

**Block 1 comparison from Studies 1 and 2: Results and discussion.** Results from a 3 (Block Instructions) x 2 (Prime) x 2 (Target) MLM established a significant
main effect for prime ($F(1, 2403) = 8.38, p = .004$) and a marginal effect of target ($F(1, 2403) = 3.67, p = .056$), suggesting $P_E$ amplitude was greater during white compared to black prime trials, and gun (marginally) compared to tool trials (Figure S1b). This qualified the Prime x Target interaction ($F(1, 2403) = 72.25, p < .001$), whereby $P_E$ was larger during white-gun trials ($M = 5.08$), followed by black-tool trials ($M = 4.75$). Like Study 1, these stereotype-incongruent trials yielded more positive $P_E$ signals than both types of congruent trials (black-gun: $M = 3.88$, $t(2403) = 7.97, p < .001$, and $t(2403) = 4.64, p < .001$, for white-gun and black-tool respectively; and white-tool: $M = 3.98$, $t(2403) = 7.34, p < .001$, and $t(2403) = 4.05, p < .001$; Figure S1b). There is also a Target x Instructions interaction ($F(2, 2403) = 7.41, p = .001$), which is driven by the greater divergence in target type observed during reappraise ($\Delta = 0.76, t(2403) = 3.84, p < .001$) relative to suppress ($\Delta = 0.32; |t| < 2, p > .1$) and attend ($\Delta = 0.17; |t| < 2, p > .2$; Figure S1b). Additionally, we found no differences between block instructions by testing the effect of block in a one-way ANOVA at each level of target type ($Fs < 1, ps > .4$). The analysis also revealed an interaction of Prime x Instructions ($F(2, 2403) = 3.04, p = .048$), suggesting that white primes produced an increased $P_E$ signal compared to black primes during reappraise ($\Delta = 0.48, t(2403) = 2.46, p = .014$) and suppress ($\Delta = 0.46, t(2403) = 2.35, p = .019$), whereas there was no difference during attend ($\Delta = 0.03; |t| < 1, p > .8$; Figure S1b). Again, no differences between block instructions were found by testing the effect of block in a one-way ANOVA at each level of prime type ($Fs < 1, ps > .4$). Unlike Study 1, here we found no three-way interaction of instructions, prime, and target ($F < 1, p > .6$).
Although this was not one of the primary aims of the current experiments, these exploratory results have implications for the debate on the functional significance of the P_E. Given how the P_E results diverge from those reported of the ERN, the present findings support previous research (Overbeek et al., 2005) indicating that the P_E is distinct from the ERN.

Interestingly, the Prime x Target interaction (evident in both within- and between-subjects comparisons) seems to provide evidence in favor of the motivational significance hypothesis. Here, we found that the “incongruent” trials elicited a more positive P_E than the “congruent” trial types. For the WIT, errors made in response to incongruent trials (i.e., those with greater response conflict) would perhaps have more motivational significance than congruent trials in terms of control implementation. Although this sounds similar to how some have explained ERN results, it is important to note that the P_E pattern described here lacks the crucial affective element found in the ERN (i.e., distress). Specifically, while completing the task sans the application of emotion regulation strategies, P_E does not appear to distinguish between “incongruent” trial types, in spite of the fact that committing errors during one of these trial types in particular (i.e., black-tool) is potentially more distressing than the other (i.e., white-gun).

Recently, some researchers have suggested that the P_E is comparable an ERP component known as the P300, which is thought to reflect the motivational salience of stimuli (e.g., Ridderinkhof et al., 2009; Davies et al., 2001). In line with this functional explanation for the P300, it is generally found to be greater in response to novel stimuli, and recent research suggests P_E can be associated with infrequent errors (Ridderinkhof et al., 2009; Davies et al., 2001). However, in the present studies, the amplitude of P_E is not
enhanced for more infrequent error types. For example, black-tool was generally the condition during which participants made the greatest percentage of errors on average, but yet it also generated a larger $P_E$ amplitude relative to other conditions.

It is more difficult to interpret the impact of the emotion regulation strategies on the $P_E$. As mentioned in the main manuscript, the sample size for the between-subjects analysis was most likely too small to uncover the block differences observed previously in the within-subjects analysis. Regardless, we did find differences in the main effects of both prime and target within the blocks. During reappraise, $P_E$ was more positive in response to guns than tools, whereas there was no main effect of target during other block instructions. Both reappraise and suppress yielded larger $P_E$ amplitudes when the prime was white compared to black. These results are difficult to interpret, and do not fit easily with the extant literature the functionality of the $P_E$.

Further research is necessary to clarify the relationship between the $P_E$ component and error-related affect before any meaning can be derived from the present results. Previous work has found that the amplitude of $P_E$ decreases as the negative affect produced by making errors increases (Hajcak, McDonald, & Simons, 2004). However, this direct link between $P_E$ and negative affect yet to be reproduced in subsequent research (Overbeek et al., 2005). It is perhaps premature to investigate the impact of attempting to attenuate emotional responses to errors on $P_E$ when so little is known about how $P_E$ is influenced by error-related affect.
References


Supplementary Figures

(a) $P_E$ during Study 1 (within-subjects comparison)

(b) $P_E$ during Block 1 (between-subjects comparison)

Figure S1. Comparison of $P_E$ across blocks during (a) Study 1 (within-subjects) and (b) the first block during Studies 1 and 2 (between-subjects). Displayed are the composites of response-locked, weighted grand averages from nine 3 x 3 electrodes centered on Cz (i.e., FC1, FC2, FCz, C1, C2, Cz, CP1, CP2, and CPz). Orange highlighting denotes the quantification window for $P_E$ (150 - 250 ms post-error response). $P_E$ = error positivity.
Figure S2. The effect of block order on suppress (a and b) and reappraise (c and d). (a & c) Accuracy rates during the WIT (displayed on the y-axis) as a function of block order and instructions. (b) For the PDP data, a comparison of suppress with attend during blocks 1 and 2 yielded a significant interaction of Block Instructions x Order Group x Process. (d) PDP analysis of reappraise versus attend during blocks 1 and 2 revealed an
interaction of Block Instructions x Block Order x Prime. PDP = process dissociation procedure.
Figure S3. The effect of block order on the ERN. (a) The ERN is displayed for both suppress and attend when they occurred as both the first and second blocks. (b) Comparing the ERN during reappraise and attend during the first and second block. ERN = error-related negativity.