

Does it Bug You? Exploring the Role of Self-Affirmation in the Distress Caused by Killing

by

Jamin Blatter

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Psychology

University of Alberta

© Jamin Blatter, 2017

## Abstract

A large body of research confirms that self-affirmation reduces defensive reactions to self-threats. Self-affirmation buffers against threats to the self by bolstering perceptions of the self as moral and good, thereby serving to restore global self-integrity. However, recent research shows that when one affirms on a domain related to the initial threat, self-affirmation can backfire, as it serves as a reminder of the self-standard that had just been violated. A series of studies examined the effect of self-affirmation on a unique type of self-threat: killing. Because killing likely represents a stark departure from one's moral standards, it was hypothesized that self-affirmation may draw attention to the violated standard, and thus fail to bolster one from distress, and may even exacerbate it. I tested this hypothesis across four studies designed to assess the effect of self-affirmation on guilt and distress after immoral behavior, namely killing bugs. I then assessed psychological distress using various indices of guilt and shame, including a behavioral hand washing measure. Contrary to the hypothesis, Study 1 found that when given the choice of how many bugs to kill, self-affirmation might buffer against the distress caused by killing, as self-affirmed participants reported less distress than non-affirmed participants. However, a different picture emerged in subsequent studies, such that self-affirmation led to more guilt and distress when participants were forced to kill a specified number of bugs, in line with the hypothesis (Studies 2 and 4), and did not seem to help or harm participants if self-affirmation was completed after the extermination task, rather than before (Study 3). These results suggest that the conditions under which the immoral behavior occurs as well as the timing of the self-affirmation intervention may influence the level of distress experienced. Discussion focuses on the theoretical implications and importance of these findings within the greater self-affirmation and distress literatures.

## **Preface**

This thesis is an original work by Jamin Blatter. The research project, of which this thesis is a part, received research ethics approval from the University of Alberta Research Ethics Board, “Examining the Role of Self-Affirmation as a Buffer Against Killing,” Pro00050780, October 10, 2014.

## Acknowledgements

I would like to extend a sincere and deep thanks to my supervisor, Jeff Schimel, not only for his invaluable support and guidance throughout this project, but throughout my graduate career. I would not have made it to the finish line without his help and advice throughout the process of conducting and reporting the research presented in this dissertation. I am indebted to you. I am also eternally grateful to my colleagues (and friends) in the lab: David Webber for starting me on this journey down the rabbit hole by introducing me to this research and paradigm, and for showing me how it's done; Erik Faucher for helping me prepare for my defense by showing me just how little I knew (about anything), and motivating me to stretch my intellectual expertise; Mike Sharp for supporting (and tolerating) me when I needed the extra motivation and could no longer see the light at the end of the tunnel; as well as other honorary members and friends of the lab who each contributed in various ways to the end product of this dissertation, including the newest member of the lab, Andy Scott.

Additionally, the research reported herein could not have been completed without the extremely diligent and unwavering hard work of a team of gifted research assistants, who dealt with a sensitive paradigm with the grace and wisdom of true scientists: Tanya Pacholok, Kerry Howell, Tara Sunde, Heather Schultz, Stephanie Mah, Teagan O'Connor, and Sarah Van Kuppeveld. I cannot thank you enough. I would also like to thank the members of my supervisory committee, and recognize the invaluable insight and feedback they provided. Many thanks to Jenn Passey, David Rast, Elena Nicoladis, Sarah Moore, and Toni Schmader.

Last, but certainly not least, I would like to thank my family and friends, particularly my parents, John and Bev Blatter, for their unfailing faith in me, and for supporting me throughout this (extremely) long process, especially when I did not have faith in myself.

## Table of Contents

General Introduction .....	1
Cognitive Dissonance Theory.....	2
Self-Affirmation Theory.....	4
Self-Affirmation (is Good?).....	7
Self-Affirmation (is Bad?).....	8
Guilt and Shame.....	12
The Problem with Killing.....	14
The Present Research.....	17
Study 1 .....	20
Method .....	21
Participants.....	21
Research Design.....	22
Materials and Experimental Manipulations.....	22
Procedure .....	27
Results.....	31
Manipulation Checks .....	31
Gender Effects .....	32
Effect of Condition on Supplemental Measures (Self-Esteem and Affect).....	32
Effect of Condition on Self-Report Distress.....	33
Effect of Condition on Willingness to Kill and Number Killed.....	35
Effect of Condition on Behavioral Guilt (Self-Punishment) .....	35
Discussion.....	35

Study 2 .....	39
Method .....	40
Participants.....	40
Research Design.....	41
Materials and Procedure .....	41
Results.....	43
Manipulation Checks .....	43
Gender Effects .....	44
Effect of Condition on Supplemental Measures (Self-Esteem and Affect).....	45
Effect of Condition on Self-Report Distress.....	46
Effect of Condition on Behavioral Guilt (Hand Washing) .....	47
Discussion.....	49
Study 3 .....	53
Method .....	53
Participants.....	53
Research Design.....	54
Materials and Procedure .....	55
Results.....	57
Manipulation Checks .....	57
Gender Effects .....	57
Effect of Condition on Supplemental Measures (Self-Esteem and Affect).....	58
Effect of Condition on Self-Report Distress.....	59
Effect of Condition on Behavioral Guilt (Hand Washing) .....	61

Discussion.....	64
Study 4 .....	67
Method.....	68
Participants.....	68
Research Design.....	69
Materials and Procedure .....	69
Results.....	71
Manipulation Checks .....	71
Gender Effects .....	71
Effect of Condition on Supplemental Measures (Self-Esteem and Affect).....	72
Effect of Condition on Self-Report Distress.....	73
Effect of Condition on Behavioral Guilt (Hand Washing).....	76
Discussion.....	79
General Discussion .....	83
Addressing Inconsistent Findings across Studies .....	84
Limitations .....	85
Gender.....	85
Exclusions.....	86
No Kill Control Conditions.....	86
Generalizability of Findings .....	87
Framing the Findings within the Self-Affirmation and Distress Literatures .....	90
More than Self-Discrepancy? .....	90
Timing of Self-Affirmation.....	92

Implications.....	93
Conclusion .....	97
References.....	99

## List of Tables

<b>Table 1:</b> Means, Standard Deviations, Cell Size, and Test Statistics for Each Individual Distress Variable from ANOVA on Self-Report Distress Composite, Including Total Scores for TRGI and SSGS (Study 1).....	34
<b>Table 2:</b> Means, Standard Deviations, Cell Size, and Test Statistics for Each Individual Distress Variable from ANOVA on Self-Report Distress Composite, Including Total Scores for TRGI and SSGS (Study 2).....	47
<b>Table 3:</b> Means, Standard Deviations, Cell Size, and Test Statistics for Each Individual Distress Variable from ANOVA on Behavioral Distress Composite, Including Both Raw and Standardized Z Scores for Time, Towels, and Soap (Study 2).....	48
<b>Table 4:</b> Means, Standard Deviations, Cell Size, and Test Statistics for Each Individual Distress Variable from ANOVA on Self-Report Distress Composite, Including Total Scores for TRGI and SSGS (Study 3).....	60
<b>Table 5:</b> Means, Standard Deviations, Cell Size, and Test Statistics for Each Individual Distress Variable from ANOVA on Behavioral Distress Composite, Including Both Raw and Standardized Z Scores for Time, Towels, and Soap (Study 3).....	63
<b>Table 6:</b> Planned Contrasts of Kill/Affirmation, Kill/No Affirmation, and No Kill/No Affirmation Conditions on All Behavioral Distress Measures (Study 3).....	63
<b>Table 7:</b> Means, Standard Deviations, Cell Size, and Test Statistics for Each Individual Distress Variable from ANOVA on Self-Report Distress Composite, Including Total Scores for TRGI and SSGS (Study 4).....	75
<b>Table 8:</b> Planned Contrasts of Kill/Affirmation, Kill/No Affirmation, and No Kill/No Affirmation Conditions on All Self-Report Distress Measures (Study 4).....	76

**Table 9:** Means, Standard Deviations, Cell Size, and Test Statistics for Each Individual Distress Variable from ANOVA on Behavioral Distress Composite, Including Both Raw and Standardized Z Scores for Time, Towels, and Soap (Study 4)..... 78

**Table 10:** Planned Contrasts of Kill/Affirmation, Kill/No Affirmation, and No Kill/No Affirmation Conditions on All Behavioral Distress Measures (Study 4)..... 78

## General Introduction

“We are reluctant to admit that essentially war is the business of killing...while the soldier himself comes from a civilization in which aggression, connected with the taking of life, is prohibited and unacceptable” (Baum, 2004, p. 2).

On a daily basis, people often encounter situations in which their behavior runs counter to their perceptions of themselves. Individuals may say something they wish they had not (e.g., agree to an opinion contradicting their true feelings), act in a way that recklessly puts either themselves or another at risk or causes harm, or they may simply underperform in a valued domain such as school or work (Aronson, Blanton, & Cooper, 1995). Moreover, most people do not enjoy receiving threatening feedback suggesting their actions have been harmful to themselves or others, that their actions contradict their values or beliefs, or that their beliefs and values are incorrect (Crocker, Niiya, & Mischkowski, 2008). Even though this type of feedback can often be informative and useful for correcting one’s behavior, it also poses a threat to one’s image of oneself as good, moral, intelligent, rational, responsible, caring, and healthy (Crocker et al., 2008).

In other words, on a day-to-day basis, the opportunity to encounter events or information posing a threat to one’s sense of self as moral and competent (i.e., one’s overall sense of self-integrity) seems to be without bounds (Sherman & Cohen, 2006). Because this type of feedback is unpleasant, individuals tend to react to threatening information in a defensive manner. Researchers have therefore identified a number of strategies people invoke to manage the threat unflattering information poses to self-integrity, allowing them to preserve a positive self-image (Critcher, Dunning, & Armor, 2010).

## **Cognitive Dissonance Theory**

One theory useful for examining the threat caused by a mismatch between one's self-views and actions is cognitive dissonance theory (Festinger, 1957). Cognitive dissonance (or simply dissonance) theory suggests people experience a state of negative psychological arousal or tension when two beliefs, thoughts, or actions (referred to as "cognitions") are accessible at a given moment that are inconsistent with each other. For example, dissonance may occur when one realizes one has behaved (with little reason or justification) in a manner that runs contrary to one's beliefs or attitudes. In other words, if you hold a specific thought or belief and do something that does not align with that belief (or is inconsistent with your values), you experience dissonance (Festinger, 1957).

The classic example of a situation likely to arouse dissonance (which Festinger used to illustrate his theory) relates to smoking. Information regarding the ill effects of smoking is widespread, yet many people continue to smoke. Thus, the fact that smoking is harmful and the knowledge that one continues to smoke in spite of this, is inconsistent and should therefore provoke dissonance (Festinger, 1957). Killing seems to be a clear example of a situation that might also elicit powerful dissonance, as most people "know" (and believe) killing is morally wrong, yet their actions do not align with this knowledge. Indeed, some researchers have argued killing is likely to arouse feelings of distress, such as guilt, as it "conveys a violation or departure from the standards of behavior" (Lee, Scragg, & Turner, 2001, p. 461). The act of taking a life is a violation of personal as well as societal standards of what is right and moral, creating a discrepancy between one's behavior and beliefs, leading to feelings of dissonance or distress.

Moreover, according to the theory, the dissonance caused by inconsistency is unpleasant. Individuals are therefore motivated to reduce the discrepancy and achieve "consonance" between

their cognitions, in order to attain psychological equanimity (Festinger, 1957). One strategy that allows people to combat the dissonance aroused by self-threats involves rationalization or self-justification in some manner, such as changing beliefs or attitudes in regards to the threatening information. Dissonance researchers (e.g., Brehm & Cohen, 1962; Cooper & Fazio, 1984; Festinger, 1957; Wicklund & Brehm, 1976) have studied this type of rationalization or self-justification extensively.

Indeed, a great deal of research supports the notions advanced by Festinger and dissonance theory, although several revisions to the original conceptualization have been proposed over the last several decades since its inception (e.g., Aronson, 1968; Aronson, 1969; Cooper, 1971; Cooper & Brehm, 1971; Cooper & Fazio, 1984; Cooper & Worchel, 1970; Nel, Helmreich, & Aronson, 1969). Many of these revisions extended the theory as well as attempted to outline the boundaries under which inconsistencies in cognitions cause dissonance and subsequent changes in attitudes and behavior. Aronson (1968), for example, suggested only threatening an important aspect of one's self-concept would lead to dissonance. Researchers showed that when participants wrote counterattitudinal essays in support of the legalization of marijuana, this led to the expected change in attitudes (or attempts to reduce dissonance), but only when participants' morals or values were challenged (Nel et al., 1969). Other researchers argued it might be necessary for one to feel a sense of personal responsibility for negative outcomes (or causing harm) in order to arouse dissonance (Cooper & Fazio, 1984). Still others have proposed that in order for inconsistency (or counterattitudinal behavior) to evoke dissonance, the action leading to negative consequences must be freely and knowingly chosen (e.g., Cooper, 1971; Cooper & Brehm, 1971; Cooper & Worchel, 1970).

In spite of these proposed revisions to the theory, most researchers appear to concede that once dissonance is aroused, people are motivated to reduce it, as originally suggested by Festinger (1957). Importantly, researchers have argued that in order for such efforts to effectively address and reduce the dissonance, one must direct one's actions at the initial behavior that aroused the dissonance (Blanton, Cooper, & Aronson, 1997). Accordingly, when participants in dissonance (e.g., forced-compliance) studies reported changing their attitudes or behavior, this indicated the participants were trying to account for or undo the initial act. In each instance, after individuals addressed the original action that first aroused the dissonance, through changing their attitudes or otherwise engaging in self-justification, the need to reduce the dissonance dissipates (Blanton et al., 1997).

### **Self-Affirmation Theory**

According to Festinger (1957) and dissonance theory, the desire to reduce dissonance appears to originate from a need to remain psychologically consistent, particularly concerning cognitions deemed important to the self (Steele & Liu, 1983). The mere existence of inconsistent cognitions is enough to evoke dissonance and thereby initiate processes aimed at reducing this negative arousal (Festinger, 1957). Although a vast literature demonstrates a need to reduce dissonance in the face of behaving inconsistently with one's beliefs or values, a recent revision to the theory suggests that a threat to the self (or ego) is responsible for dissonance, rather than mere psychological inconsistency (Steele & Liu, 1983).

Steele and Liu (1983) argue humans have a need to maintain a positive self-image, and when people behave inconsistently with important beliefs or values, these actions threaten that positive self-image, causing dissonance. Steele and Liu (1983) therefore proposed self-affirmation theory as an alternative to dissonance theory, suggesting dissonance processes arise

out of a need to maintain positive self-regard, rather than out of a need for cognitive consistency. Individuals have a deep desire to avoid looking (and importantly, feeling) foolish or incompetent. When presented with information that threatens this self-concept, a need to maintain a positive self-view causes dissonance and feelings of distress (Steele & Liu, 1983).

According to Steele (1988), the research on dissonance theory had typically confounded inconsistency in cognitions with threats to the self and self-integrity. Steele (1988) argued individuals are actually motivated to reduce dissonance because of the negative implications inconsistency has for the self, rather than due merely to inconsistency per se. Defensive responses are not the result of inconsistencies related to causing harm to others, making a difficult choice, or giving a speech contrary to one's beliefs. Instead, self-affirmation theory maintains the self-defensive responses occur because the individual's behavior indicates the individual is inept, unintelligent, or immoral (Aronson et al., 1995; Nel et al., 1969; Scher & Cooper, 1980; Steele, 1975).

These ideas prompted important insight into potential alternatives to reducing dissonance, rather than the usual change in beliefs or attitudes. If threatening one's self-image or self-integrity does indeed cause dissonance, then one need not necessarily react in the ways typically outlined in the original dissonance research. Rather, it may be possible to restore the self-concept or one's self-image by affirming a cherished aspect of the self that may be unrelated to the initial threat (Aronson et al., 1995). To continue the example discussed previously, smokers do not necessarily have to rationalize their harmful behavior (e.g., claiming smoking is not that harmful) to reduce dissonance, if the person instead affirms the self in another valued domain (e.g., by reminding him or herself that he/she is a good friend, student, or spouse).

According to self-affirmation theory, humans are fundamentally motivated to preserve self-integrity, meaning people are motivated to maintain an image of themselves as moral and good (Cohen & Sherman, 2007). However, what happens when one's belief about the self (and self-integrity, in particular) is threatened? Self-affirmation theory examines the process of how people strive to maintain the overall integrity of the self in the face of such threats (Sherman & Cohen, 2007). That is, according to self-affirmation theory, although individuals may want to uphold distinct images pertaining to the self, such as viewing oneself as a "good worker" or "good Christian," this is not the principal goal. People instead strive to sustain faith in their overall or global self-integrity, a general conception of one's virtue or goodness. Thus, there may be a great deal of flexibility in terms of the domains in which individuals base their self-integrity, and, importantly, self-affirmation theory proposes individuals may be better able to manage a threat to the self in one domain if they have other domains in which to feel good about themselves (Sherman & Cohen, 2007). These types of affirmations allow one to bypass typical defensive responses to threats to the self, thereby eliminating dissonance and other negative emotions.

Self-affirmation then is defined as the restoring or bolstering of the "perceived integrity of the self, its overall adaptive and moral adequacy" (Steele, 1988, p. 291) through cognitive or behavioral means. Psychological interventions that allow one to self-affirm (and therefore restore self-integrity) can encompass a wide range of events, including (most often) reflecting on and describing values or characteristics one finds personally important (e.g., positive aspects of the self), or receiving information about oneself that is favorable (Schmeichel & Vohs, 2009). One of the simplest (yet still powerful) self-affirmation interventions involves ranking one's core values and characteristics in terms of personal importance. This simple ranking task seems to

serve as a reminder of one's positive attributes. Thus, although one may have received a threat to one area of the self-concept, completing this task reminds the person that he or she still has other values and characteristics that are important for defining the self. This type of affirmation serves to restore the overall sense of the self as moral and worthwhile in spite of negative feedback.

### **Self-Affirmation (is Good?)**

According to self-affirmation theory, receiving a threat to one's global sense of self (or self-integrity) evokes dissonance. However, unlike the other conceptualizations of dissonance, self-affirmation theory proposes that it is possible to reduce dissonance without directly dealing with the initial action that aroused the dissonance. Researchers have stated, "because the disturbing thing about dissonant behavior is its ego threat, any self-affirming activity may reduce dissonance even when it does not resolve or dismiss the particular provoking inconsistency" (Steele & Liu, 1983, p. 18). Steele and Liu (1983) therefore suggest prior research, employing the forced-compliance procedure, may have obtained the results detailing self-justifying responses to dissonance only because these researchers had not allowed participants to react in other ways. Thus, participants may not have chosen to alter their beliefs or attitudes if they had instead been provided an opportunity to engage in self-affirmation (Blanton et al., 1997).

Steele and Liu (1983) demonstrated support for the proposition that self-affirmation can mitigate the effects of dissonance using the same forced-compliance paradigm used in most dissonance research, in which participants were asked to write essays in favor of steep tuition hikes (which runs counter to most students' beliefs). However, the researchers altered the paradigm, such that, in certain conditions, participants engaged in self-affirmation (or not), prior to advocating for the tuition increase. Importantly, results showed participants did not

subsequently change their attitudes in favor of raising tuition in the high-dissonance condition, if they first engaged in self-affirmation of a personally important (but unrelated) value (Blanton et al., 1997; Steele & Liu, 1983).

Since this initial work by Steele and Liu (1983), a substantial body of research supports the notion that affirming the self in an unrelated domain reduces a wide variety of defensive responses to self-threats (Cohen & Sherman, 2007; see Sherman & Cohen, 2006 for a review). For example, affirming the self lowers defensive evaluations of arguments that oppose one's ideology (Cohen, Aronson, & Steele, 2000), makes one more accepting of information suggesting susceptibility to health risks (e.g., Harris & Napper, 2005; Reed & Aspinwall, 1998), and causes one to engage in prosocial behaviors, even if helping another person succeed is threatening (Tesser, Martin, & Cornell, 1996). Additionally, self-affirmation benefits diverse domains, including reducing stress following a mentally taxing task (Keough, Garcia, & Steele, 1998), spending less time ruminating after failure (Koole, Smeets, van Knippenberg, & Dijksterhuis, 1999), eliminating prejudiced attitudes (Shrira & Martin, 2005), counteracting stereotype threat (Martens, Johns, Greenberg, & Schimel, 2006), restoring self-control resources (Schmeichel & Vohs, 2009), and eliminating worldview defense after a mortality reminder (Schmeichel & Martens, 2005).

### **Self-Affirmation (is Bad?)**

The research demonstrating support for the efficacy of self-affirmation for ameliorating defensive reactions to self-threats across a wide range of domains is compelling. However, it is also interesting to note that other procedures designed to engage self-focus of a similar nature can sometimes backfire, leading to the reverse outcome (Aronson et al., 1995). Research on self-awareness (e.g., Duval & Wicklund, 1972; Gibbons, 1990; Scheier & Carver, 1977), for

example, shows that individuals do not particularly enjoy self-focus after behaving in a dissonant manner, as it is unpleasant and uncomfortable (Greenberg & Musham, 1981). Self-awareness (self-focus) has been shown to increase negative emotions (Gibbons, 1987) and heighten ego defensive reactions (Federoff & Harvey, 1977). Research also indicates that focusing on positive self-standards can elicit negative affect, perhaps because these standards contrast with one's current views of self (i.e., current self-appraisals). Thus, if one's self-view does not match a particular self-standard, this discrepancy is likely to arouse a host of negative emotions, such as dejection or agitation (Higgins, 1987).

What this suggests is focusing on the self may lead to one of two contrasting outcomes. On the one hand, research clearly indicates self-focus may produce a self-affirmation effect, reducing the tendency to respond defensively to self-threats by restoring a positive view of oneself. On the other hand, focusing on the self may create a self-discrepancy effect, increasing the tendency for defensive responses by creating a negative view of oneself in relation to one's standards (Aronson et al., 1995). Aronson and colleagues (1995) argue that perhaps these opposing effects occur as a result of positive self-views serving a dual function of being both a standard from which to compare and a resource from which to draw upon. If one's self-view is threatened (e.g., by engaging in behavior that counters one's values), then focusing on that particular self-view may highlight the standard that was violated (i.e., prompting comparison and drawing attention to the fact one has fallen short), thereby increasing the perception of self-threat. Under other circumstances, that specific self-view might serve an affirmational resource function if the self-view that was threatened is in a different domain (i.e., providing alternative sources of worth). Aronson and colleagues (1995) conclude the cause of pride in one context

can be the cause of shame in another instance, depending on the source of the self-threat and subsequent self-focus.

Participants in most self-affirmation studies (e.g., Steele & Liu, 1983) are asked to “freely” choose to write an essay that counters their views (e.g., in favor of raising tuition), which threatens their self-concept and creates dissonance. However, when these participants are allowed to affirm an important value in an unrelated domain (e.g., aesthetic preferences), this effectively restores self-integrity, thereby eliminating the need to engage in defensive responding (e.g., rationalization), as dissonance is not aroused. Nevertheless, let’s imagine these participants instead affirmed on a related value (e.g., integrity or honesty). In this instance, affirming on a related domain would force the individual to focus on the fact one violated an important standard. The threat to self will stand out and the need to engage in rationalizations will remain. Aronson and colleagues (1995) therefore argue that revisiting the self-domain implicated by one’s discrepant actions will not provide any solace from the dissonance that is provoked. Conversely, affirming an aspect of the self-concept that is unrelated to the dissonant behavior will be more effective at restoring one’s self-integrity, as it does not focus on the self-standard that was threatened, and, therefore, allows the threat to dissipate (Aronson et al., 1995). Whether self-affirmation will produce a negative self-discrepancy effect or a positive self-affirmation effect will depend on the aspect of self one affirms in response to a self-threat.

Researchers tested this idea in a study, which sought to examine what would happen if individuals affirmed on a value that *was* related to the dissonance evoking action, instead of affirming a component of the self-concept that was not threatened by the individuals’ behavior (Blanton et al., 1997). To address this question, Blanton et al. (1997) asked participants to write essays (counter to their beliefs) advocating for reduced funding and support for students with

disabilities on campus (Blanton et al., 1997). In line with previous dissonance and self-affirmation research, half of the participants were led to believe they had a choice in writing the essay (high dissonance condition), whereas the other half were told they had to write the essay and did not have a choice (low dissonance condition). Once the essays were completed, participants received either relevant (compassionate) feedback, irrelevant (creative) feedback, or no feedback based on a personality test they had taken previously. Participants then rated their attitudes regarding funding for students with disabilities (to test for post-essay attitude change). In support of self-affirmation theory, participants exhibited less attitude change after affirming an aspect of the self-concept that was unrelated to the dissonant behavior. However, participants who affirmed on a domain related to the dissonant behavior later exhibited greater attitude change towards the position they had advocated (Blanton et al., 1997).

Based on these findings, affirming a relevant value one has recently violated, especially one relating to an important moral dimension (such as compassion), highlights the fact one has behaved immorally. Asking participants to reaffirm a value that was just threatened reminds them of their connection to their morals and values. In cases like this, participants are trying to maintain the view that they are good, compassionate people, who act in accordance with their morals and values through the self-affirmation exercise. Yet, they just behaved in a manner that contradicts the same morals that were self-affirmed. The discrepancy between people's beliefs and morals and the knowledge they have just behaved contradictory to those beliefs becomes more difficult (if not impossible) to reconcile (Blanton et al., 1997). In other words, when people affirm on the same self-dimension that was initially threatened, this affirmation draws attention to the discrepancy, and thus fails to reduce the dissonance.

As noted above, other research further supports the results from Aronson et al. (1995), showing there is an increase in negative emotions when we focus on positive aspects of the self, if this focus contrasts with the current self-view (Gibbons, 1987; Higgins, 1987). When one's perception of self is unfavorable (e.g., after behaving in a manner that contradicts one's values), focusing on aspects of the self related to the threatened domain may increase negative feelings about oneself. The aim of the current studies was to determine if the relation between focusing on the self and negative feelings applies to a deed that should create a negative self-image for most people, namely, killing. Because killing deeply violates moral standards, it should serve as a pervasive threat to one's image of oneself as moral and good. As such, self-focus of any kind (including self-affirmation) may fail to resolve the dissonance and distress that such an immoral act is likely to arouse, and may indeed aggravate one's psychological distress.

### **Guilt and Shame**

The evidence presented so far suggests dissonance should be aroused when one's behavior does not align with one's beliefs, self-perceptions, or self-standards. The current studies endeavor to take these ideas one step further, arguing shame and guilt are likely to be elicited in response to performing an act that is not only dissonant with one's values, but that is deemed immoral. It is possible participants are feeling more than just a state of negative arousal (i.e., dissonance) after reaffirming a relevant standard they had just violated. Perhaps one may even feel regret, remorse, guilt, or shame at the thought of one's recent actions (that are so contradictory to one's values). Thus, one would expect measures of dissonance (or distress) to detect these specific negative feelings.

According to Tangney and colleagues, the moral emotions of shame and guilt follow from a discrepancy between one's characteristics or behavior and one's moral standards

(Tangney, Niedenthal, Covert, & Barlow, 1998). Although shame and guilt are distinct “moral emotions,” many researchers acknowledge the two are often confused or used interchangeably in practice (e.g., Lewis, 1971; Tangney et al., 1998). One distinction proposes that guilt tends to refer to feelings evoked by focusing on a particular action or transgression, whereas shame tends to refer to feelings evoked by focusing on the self as a whole (Tangney et al., 1998). In other words, guilt generally refers to negative feelings pertaining to a particular action or event, which is separate from the overall sense of self. Feeling guilty typically entails perceptions of regret, remorse, or tension after doing something “bad.” Conversely, shame generally refers to negative feelings pertaining to a global self-evaluation. Feeling shame entails a judgment in which one perceives the self to have fallen short of the standards to which is it compared. Shame can be a very powerful (and painful) emotion, and is often associated with feeling “small,” wanting to shrink, disappear, or escape from others (real or imagined) to whom one feels exposed (Tangney et al., 1998).

Although this distinction between shame and guilt is important for many researchers on theoretical grounds (e.g., Higgins, 1987), for the present purposes, I expected violating moral standards by killing would elicit both shame (an evaluation of the self as immoral) and guilt (feeling badly about the specific act), and both facets were therefore examined in the present studies. In fact, when outlining the importance of dealing carefully with veterans returning from war, Maguen et al. (2010) noted that these individuals “who have killed may experience significant shame and/or guilt...” and this awareness should factor into any therapy provided (Maguen et al., 2010, p. 90), suggesting both emotions are likely present after killing.

## **The Problem with Killing**

As children, most people learn that aggression is morally unacceptable. Moreover, people learn that killing (especially another human) in particular is the absolute immoral, illegal, and universally wrong act. This lesson is usually instilled by one's caregivers, and is further reflected by society with harsh penalties should one disobey this moral standard (Hagen, 1994). The act of taking a life is a violation of one's deep-rooted morals and values that the majority of individuals endorse through socialization (Bandura, 1990). Killing is therefore particularly likely to evoke dissonance. Additionally, because the act of killing represents such a clear violation from standards of morality, strategies aimed at reducing this dissonance through bolstering perceptions of the self as good and moral may fail. In fact, affirming the sense of oneself as decent and moral may actually highlight the discrepancy between one's moral beliefs and standards and their actual behavior (i.e., killing), perhaps even exacerbating the feelings of distress. The result of these attempts to restore self-integrity may actually backfire, creating a high level of dissonance and distress that may prove difficult to counteract (Webber, Schimel, Martens, Hayes, & Faucher, 2013).

Although most cultures uphold the belief that killing is wrong, there are circumstances in which it is forgivable. For example, the act of war is certainly not a new concept, and the history of many civilizations are rife with examples of the absolute worst of human behavior, namely, extended and mass killing (Martens, Kosloff, Greenberg, Landau, & Schmader, 2007). One particularly alarming fact is that the act of killing itself may instigate further killing. Indeed, some researchers have reasoned the act of taking a life may lead to further killing because of the need to justify one's immoral actions (and possibly reduce dissonance). Researchers posit an increase in subsequent killing may result from striving to deal with the guilt one experiences, as

well as from a need to counteract the negative implications for one's character that may be triggered by the initial killing (e.g., Martens et al., 2007).

As killing is likely to violate one's self-standards regarding moral decency, it is also likely to pose a threat to the self, arousing psychological discomfort (i.e., dissonance and distress). To the extent this is true, defensive processes aimed at reducing the distress that is aroused are likely to be initiated after one has killed, thereby enabling one to cope with the threat to self (Martens et al., 2007). In other words, when a person has killed, he or she should be motivated to reduce the unpleasant emotions elicited due to behaving in a morally discrepant fashion. As discussed previously, these efforts entail attempts to minimize dissonance evoked by engaging in counterattitudinal behavior (Festinger, 1957). Importantly, dissonance theory suggests one strategy for reducing this psychological discomfort is to convince oneself the behavior was justified (Brehm & Cohen, 1962). This type of justification can cause one to maintain the discrepant behavior, or perhaps even lead to increased frequency of the behavior. This idea led some researchers to conclude (somewhat disconcertingly) that killing may be one such behavior likely affected by such efforts of justification. As such, one may further perpetuate the act of killing if one has already engaged in this immoral behavior, as part of a need to reduce the psychological threat posed by previous killing (Martens et al., 2007).

In order to examine these ideas and to assess the aftermath of direct killing, it is necessary to provide an environment in which killing is induced. Researchers have therefore utilized a bug-killing paradigm designed to simulate "real-world" killing behavior, allowing them to determine the impact initial acts of killing may have on fueling subsequent killing (e.g., Martens et al., 2007; Martens, Kosloff, & Jackson, 2010). To this end, the researchers (Martens et al., 2007) varied the number of bugs participants first had to kill as practice, and then

examined the number of bugs killed at a later point during a timed, self-paced killing task. Results revealed that when participants had to kill more bugs at the start, they subsequently killed more bugs during the self-paced extermination task. However, only participants who indicated that they perceived a degree of similarity between themselves and small insects showed this effect (Martens et al., 2007). This increase in killing during the second phase of the extermination task appeared to result in positive emotional change. According to Martens et al. (2007), this change reflects an attempt to reduce the negative emotions by justifying one's actions, which fuels more killing (Martens et al., 2007).

Further research using the bug-killing paradigm delineated whether this increase in killing was due merely to the effect of first practicing the killing task or specifically because of the initial act of killing (Martens et al., 2010). The researchers also wanted to determine whether performing the task serves to desensitize participants (which leads to an increase in killing), or if there is indeed a motivation to increase killing (perhaps to reduce the negative emotions aroused by feeling guilty). Results of these studies revealed that the observed increase in killing is not due to practice effects, as participants only killed more bugs when they believed they were actually killing (rather than when they knew the bugs were safe). Additionally, participants who believed they actually killed the bugs, and who killed more bugs initially, continued to exterminate more bugs during the second phase of the study. This finding implicates an underlying motivational component, rather than the suggested desensitization explanation (Martens et al., 2010). The results of these studies provide converging support for the notion that killing does indeed beget killing, suggesting this increase in killing may stem from the need to deal with the negative emotions aroused by morally discrepant behavior.

These procedures, though obviously not without their limitations in terms of the generalizability of the findings to real-world killing behavior, provide invaluable evidence and shed light on potential implications for individuals who kill (e.g., during war). These studies provide a direct simulation of actual killing behavior, where other experimental procedures examining aggression may fall short (Martens et al., 2007). Other experimental paradigms involving typical laboratory manipulations have been limited in their ability to examine killing behavior systematically and directly. Previous research procedures used to investigate other forms of aggression may provide insight into the underlying psychological processes involved in killing (as killing is perhaps the ultimate form of aggression), however, researchers propose the act of taking a life is distinct from other aggressive behaviors in terms of its permanence and impact (Martens et al., 2007).

Specifically, once one has killed, the harm is irrevocable. The individual's life cannot be restored and amends cannot be made for the loss experienced by loved ones of the deceased (Martens et al., 2007). These factors suggest killing may be especially likely to elicit feelings of distress (e.g., guilt and shame). As such, this bug-killing paradigm seemed particularly useful for the present studies, as it is a very direct measure of actual killing behavior that may induce feelings of distress. Additionally, although the "victims" of the extermination task are mere insects, participants report elevated levels of guilt and distress when they believe they are harming another animal (Webber et al., 2013). The present studies therefore used the bug-killing paradigm as an accurate simulation of real-world killing behavior.

### **The Present Research**

The evidence reviewed suggests the effect of self-affirmation may depend upon the aspect of the self being affirmed. In particular, there appears to be two sides to the research on

self-affirmation theory: demonstrating self-affirmations can either minimize or maximize the gap caused by behavior that is discrepant from one's view of self, depending on the type of affirmation. Most of the research shows self-affirmation can be effective for minimizing defensive responses and reducing dissonance after a threat. However, later research (e.g., Blanton et al., 1997) proposed that the buffering effect typically observed after self-affirming is most likely to occur when one affirms the self in a domain that is unrelated to the initial threat. If one affirms on the same dimension that was threatened, the affirmation process may backfire, leading to increased dissonance and defensive outcomes. Although several studies were reviewed that demonstrated this effect, the idea that self-affirmation can backfire and lead to detrimental effects under some circumstances is still relatively novel and raises a number of questions. For example, are there other situations (other than affirming on the threatened domain) in which engaging in self-affirmation may backfire?

Moreover, the literature on self-affirmation up to this point has primarily focused on domains of counterattitudinal behavior that do not implicate deeply held moral values (e.g., writing an essay that contradicts one's beliefs). As such, the effect self-affirmation may have on completing an act that powerfully diverges from one's morals remains largely untested. The present studies were therefore designed to determine the effect self-affirmation might have on killing, a type of threat that is unique from others in both its level of immorality as well as its permanence and impact.

As a means of examining this hypothesis, the current set of studies were designed to approximate a real-life killing experience in a laboratory context through the use of a bug-killing procedure adapted from previous researchers (e.g., Martens et al., 2007; Martens et al., 2010; Webber et al., 2013). For this task, participants adopted the role of "exterminator" and took part

in a bug-killing paradigm. Although it is well understood that the simulation of killing bugs in a laboratory setting cannot be fully equated with the experience of killing other people, this procedure has advantages over investigating killing in real-world settings as it allows for the manipulation of variables (e.g., self-affirmation) in the killing context that would not otherwise be possible. Moreover, prior research (e.g., Webber et al., 2013) has noted the extermination task does seem to mirror the real-life experience of killing (e.g., for soldiers at war) in a number of ways. First, the killing is endorsed or condoned by an authority (e.g., the experimenter); second, the target is deemed less than human (e.g., bugs are indeed not human); and third, the killing is formulaic and utilizes indirect or metaphorical terms (i.e., participants were instructed to “deposit” the bugs into the “extermination machine”; to “exterminate” the insects, rather than “kill” them; Webber et al., 2013). As such, this paradigm is effective in examining the variables of interest and serving as an approximation of what one might experience after killing, albeit killing non-humans.

I therefore adopted this bug-killing paradigm to examine the influence of self-affirmation on the distress evoked by killing, thereby adding to the growing literature in this area. Prior studies have examined the effects of initial killing on further killing (Martens & Kosloff, 2012; Martens et al., 2007, Martens et al., 2010), how social validation or consensus affects distress after killing (Webber et al., 2013), and how scapegoating and/or anthropomorphizing the insects affects moral disengagement processes in the decision to kill (Webber, 2015). In line with previous research, the aim of the present studies was to extend these findings to a new research domain: self-affirmation.

This bug-killing paradigm is particularly useful, as the purpose of the current studies was to create a discrepancy between participants’ self-concepts and actual actions. Participants may

generally believe they are decent, moral people; however, completing the bug-killing task may call these beliefs into question, as killing violates widely held standards of morality. As such, the nature of the killing task was likely to arouse distress (or dissonance). As noted earlier, the lion's share of self-affirmation research suggests affirming the self might restore participants' beliefs in their "goodness," thus ameliorating the dissonance and distress they experience after killing (e.g., creating a self-affirmation effect). However, there is reason to believe when one's actions directly violate deeply held standards of morality, affirming the self may exacerbate distress because doing so highlights the inconsistency between attitudes and behavior (e.g., creating a self-discrepancy effect).

To test this idea, I conducted four studies to assess the conditions under which self-affirmation may increase the distress of killing. Study 1 provided initial evidence that self-affirmation does indeed affect feelings of distress after completing the extermination task. However, participants chose the number of bugs they wished to kill to complete the task, which led to unanticipated problems. To address the methodological issues encountered in Study 1, Study 2 forced all participants to kill a specified number of bugs, and the results further supported the hypotheses. Studies 3 and 4 were designed to build upon the results of Study 2, while also incorporating an additional condition to compare killing vs. not killing. Although the studies provided support for the overall hypothesis, a unique pattern of results also emerged for each of these studies, suggesting the timing of the self-affirmation manipulation may also affect distress levels.

### **Study 1**

The aim of Study 1 was to examine how self-affirmation processes influence the guilt and distress evoked by killing bugs. Most of the early research on self-affirmation (e.g., Steele &

Liu, 1983) demonstrates affirming the self ameliorates dissonance and distress following attitude-behavior inconsistencies. However, some work shows affirming the self may backfire (i.e., have the opposite effect) when it draws attention to the attitude-behavior inconsistency (e.g., Blanton et al., 1997). Self-affirmation bolsters the self-perception that one is a good and moral person, a perception that flies in the face of the knowledge that one has just behaved in a morally questionable manner (by killing). As such, the central hypothesis guiding this research is that self-affirmation will increase the guilt and distress of killing, a behavior that goes against widely held standards of moral conduct. To assess this hypothesis, Study 1 exposed participants to either a self-affirmation or control manipulation. Participants then completed an extermination task in which they “killed” bugs. Following this manipulation, the dependent variables of interest were several indices relating to the levels of distress (e.g., guilt and shame) participants experienced.

## **Method**

### **Participants**

Fifty-five undergraduates (22 men, 30 women, 3 unreported) from the University of Alberta were recruited in exchange for partial credit towards the completion of an introductory psychology course. Participants were at least 18 years of age ( $M = 19.37$ ,  $SD = 1.65$ ). However, nine participants were excluded due to suspicion regarding the tasks and/or research hypotheses (e.g., did not believe they were actually killing the bugs), not following instructions, or experimenter error. A final sample of 46 participants (19 men, 24 women, 3 unreported) were included in all analyses for which they had complete data ( $M_{age} = 19.49$ ,  $SD_{age} = 1.72$ ). All requirements established by the Ethics Review Board at the University of Alberta were satisfied, as well as the APA ethical standards for participation with human research subjects.

## Research Design

Study 1 used a one-way between-subjects design, manipulating one independent variable with two conditions. Participants were randomly assigned to either the self-affirmation condition ( $n = 24$ ), or no affirmation control ( $n = 22$ ) and the dependent variables included several self-report measures of distress (e.g., guilt and shame). The present study also employed a double-blind procedure in which neither the experimenter nor the participants were aware of the experimental condition to which the participants had been assigned.

## Materials and Experimental Manipulations

**Self-affirmation.** Participants in both conditions were given a list of 12 values and characteristics to rank in order of personal importance (1 = *most important*; 12 = *least important*) from the Sources of Validation Scale (Harber, 1995). The list of values included: artistic skills/aesthetic appreciation, sense of humor, romantic values, business/managerial skills, creativity, relations with friends/family, spontaneity/living life in the moment, social skills, music ability/appreciation, athletics, physical attractiveness, and neatness/tidiness (Cohen et al., 2000). After ranking the values in order of importance, participants completed a brief writing task, which served as the self-affirmation manipulation. Participants in the self-affirmation condition wrote about their highest ranked value (i.e., the value they indicated was number one) and briefly explained why the value is important to them as well as a time in their lives when it has been particularly important. Previous research (e.g., Cohen et al., 2000; Liu & Steele, 1986) has demonstrated the efficacy of this manipulation in reducing dissonance by affirming one's values, thereby serving to bolster or restore one's sense of integrity by refocusing attention on an important area of the self-concept.

Participants in the no affirmation (control) condition also ranked the list of values in order of personal importance. However, participants in this condition completed a brief writing task in which they wrote about the value they listed as being least important (i.e., the value they indicated with the number 12) and discussed why this value might be important to other people (Thomaes, Bushman, de Castro, Cohen, & Denissen, 2009). This ensured the two conditions were comparable, in that participants assigned to either self-affirmation or control were performing a writing task, however, as participants in this condition discussed a value they did not find particularly important and wrote about another person, this should not implicate the self or engage self-affirmation processes, thereby serving as a control.

**Rosenberg Self-Esteem Scale (RSES).** As an assessment of self-esteem following experimental manipulations, participants completed the RSES (Rosenberg, 1965) after the self-affirmation manipulation. This scale was included as part of a “packet of personality measures” designed to bolster the cover story (that the researchers were interested in personality and exterminators), and therefore primarily served to keep participants from focusing on the self-affirmation task. I specifically chose the RSES (as opposed to one of the myriad “personality” tests available) because it seemed to fit well with the self-affirmation manipulation (thereby avoiding arousing suspicions), is relatively simple, and can be completed quickly. This scale consists of 10 items, rated on a scale from 1 (*strongly disagree*) to 4 (*strongly agree*), with higher scores indicating higher self-esteem. For the current studies, I adjusted this scale to indicate state self-esteem, rather than trait self-esteem, as originally designed. Thus, participants indicated the way they felt at the time, as opposed to how they felt generally. Example items are “On the whole, I am satisfied with myself right now” and “I feel I do not have much to be proud of right now.”

**Trauma-Related Guilt Inventory (TRGI).** Participants completed a modified version of the TRGI (Kubany et al., 1996) as a means of examining any guilt they experienced due to completing the extermination task. This measure is designed to assess both cognitive and affective facets of traumatic guilt, drawing information from various sources including interviews with trauma survivors (e.g., war veterans and battered women), as well as both a review of the pertinent clinical literature and first-hand clinical experience. This scale consists of 32 items, rated on a 5-point scale from 1 (*not at all true*) to 5 (*extremely true*), with higher scores indicating greater levels of distress.

The original TRGI contains two factors: Global Guilt and Guilt Cognitions. The Guilt Cognitions factor contains four subscales: Wrongdoing, Lack of Justification, Distress, and Hindsight-Bias/Responsibility. Each of these four subscales represents specific factors of traumatic guilt relating to a distinct experience or event (in this case, the extermination task). Specifically, the Hindsight-Bias/Responsibility subscale evaluates one's beliefs or perceptions of personal responsibility relating to causing the event, as well as thoughts about whether or not the event could have gone differently (e.g., the changeability of the event). The Distress subscale examines negative affective feelings and arousal. The Lack of Justification subscale assesses whether there are justifiable explanations for the occurrence of the event. Finally, the Wrongdoing subscale measures whether the event contradicts moral standards regarding what is right and wrong.

The TRGI was adjusted to align with previous research (e.g., Webber et al., 2013) to more accurately reflect the experiences of the participants in the present studies and the laboratory setting. Specifically, the modified TRGI consisted of 22 items, using the same rating scheme as the original measure (i.e., using a 5-point scale). Webber et al. (2013) removed the

items on the Global Guilt scale, as this asks participants to specify the frequency with which they felt guilty since the event transpired (e.g., “Indicate how frequently you experience guilt that relates to what happened”). These items are inappropriate for the present purposes, as participants performed the bug-killing task almost immediately before completing this scale. Webber et al. (2013) also eliminated two items from the Distress subscale, as they capture how frequently the participant felt distressed since the event (e.g., “When I am reminded of the event, I have strong physical sensations such as sweating, tense muscles, dry mouth, etc.”). Webber et al. (2013) also edited the final four items from this scale to more adequately match the type of distress that was expected following the bug-killing task. For instance, the item “I feel sorrow or grief about the outcome” became “I feel troubled and concerned when I think about what happened” (Webber et al., 2013). The Justification, Wrongdoing, and Hindsight-Bias subscales were not altered, however, four items from the Guilt Cognitions scale that do not fall into any of the other subscales were also eliminated to address potential redundancy among the items.

**Positive and Negative Affect Schedule (PANAS).** To assess whether the various manipulations/tasks affected mood, participants completed the PANAS (Watson, Clark, & Tellegen, 1988). The PANAS contains 10 items regarding positive affect (e.g., *interested*, *inspired*), and 10 items reflecting negative affect (e.g., *guilty*, *hostile*). Participants indicated how they felt for each statement using a 5-point scale (1 = *very slightly or not at all*, 5 = *extremely*; Watson et al., 1988).

A variety of similar studies (e.g., Webber et al., 2013) have used the PANAS as filler items to round out the measures and to draw attention away from the rest of the questionnaires in the final packet (the TRGI and SSGS). Although the PANAS was not included with the explicit

intent to serve as another DV, it makes sense that participants might have experienced changes in mood after completing the extermination task. I therefore analyzed the PANAS in each study.

**State Shame and Guilt Scale (SSGS).** As an assessment of guilt that does not specifically relate to trauma (and one that is more typically used for research in social psychology), participants also completed the SSGS (Marschall, Sanftner, & Tangney, 1994). This scale consists of 15 items designed to assess three subscales of guilt, pride, and shame, rated on a 5-point scale (1 = *not feeling this way at all*; 5 = *feeling this way very strongly*) in which participants indicated the extent to which they felt that way presently. Items from this scale that reflect shame are “I feel worthless, powerless,” and “I want to sink into the floor and disappear.” Items reflecting pride include “I feel proud,” and “I feel good about myself” (reverse scored). Finally, some items depicting guilt are “I feel bad about something I have done” and “I feel remorse, regret.”

**Willingness to kill.** In line with previous research, a single item was included asking participants to indicate on a 9-point scale how willing they were to perform the extermination task (1 = *completely unwilling to kill*; 9 = *completely willing to kill*).

**Raffle ticket measure.** I also incorporated an additional behavioral measure of guilt to assess whether participants may have tried to reduce guilt by punishing themselves (Webber et al., 2013). Previous researchers (Nelissen & Zeelenberg, 2009) noted that although pro-social behavior may be used as a means of alleviating guilt by attempting to account for any harm one may have caused, there are certain instances in which one cannot compensate for one’s actions because the pain or damage is irreversible (such as with killing). Researchers have discovered that in these situations (in which compensatory actions cannot be made), participants may instead try to reduce their guilt by punishing themselves. I therefore utilized a raffle ticket measure to

assess self-punishment. Specifically, the researcher told participants that as a reward for completing the bug-killing task, they were eligible to enter a raffle for a gift card worth 50 dollars. The researcher then provided participants with a roll of raffle tickets and told them they could take as many tickets as they liked to enter in the raffle. I believed that if participants were experiencing feelings of guilt after the extermination task, they would engage in self-punishment for their behavior and subsequently enter the raffle fewer times, thereby punishing themselves by reducing their opportunity to win the gift card.

### **Procedure**

All participants completed the study individually. Upon arrival, the experimenter greeted participants in the hallway and brought them into the laboratory where they read and completed an informed consent form after a brief introduction to the study. As with previous research (e.g., Webber et al., 2013), a cover story was employed such that participants were told they would be completing a study investigating how exterminators deal with insects, and that this study was part of a larger research program evaluating how people in different roles deal with various animals. For this specific study, we were interested in personality and exterminators (in order to justify the use of the self-affirmation manipulations to avoid arousing suspicion). To this end, the experimenter told participants the study involved completing a bug extermination task as well as answering some questions regarding their experience. Participants then completed a series of tasks, independently and at their own pace.

All participants first completed a packet of “personality measures” (designed to bolster the cover story) containing the experimental manipulation (i.e., self-affirmation or no affirmation control) to which they had been randomly assigned, as well as the RSES (Rosenberg, 1965). Participants completed this packet in the main laboratory, which was a conference-style room.

Next, participants were ushered by the researcher from the main room to the “extermination area” (which is a separate cubicle designed to ensure privacy) in an adjacent room to complete the extermination task.

The “extermination area” contained a desk and all the materials required to complete the task. On the desk were 10 pill bugs (woodlice), a laptop with a video loaded on it, an information packet about the insects, and the extermination machine. On the left hand side of the desk were the 10 pills bugs (each measuring around 1 cm in length) each in their own clear plastic cup (which were roughly the size of a shot glass), arranged in two parallel columns of five cups each. The pill bugs belong to the family *Armadillidium vulgare*, which are a type of segmented isopods that tend to be quite small, and are gray, black, or blue in color. Woodlice are known by a variety of names, including pill bugs, potato bugs, or roly polies (as they are recognized for their defensive ability to curl into a ball). Next to the pill bugs, the extermination machine was a coffee grinder altered with a piece of tubing attached to the side, which looks as though the tube leads directly into the machine (Martens et al., 2007; Webber et al., 2013). To bolster the cover story, participants were informed the grinder would be used for the extermination task because poisonous chemicals and sprays are not allowed in the building. The right hand side of the desk had the information packet about the insects and a laptop with a video loaded on it. The experimenter instructed the participants to: (1) read through the information packet, (2) watch the video on the laptop, (3) complete the extermination task as demonstrated in the video, and (4) return to the main room for further instructions.

Prior to completing the extermination, participants were provided with some neutral information regarding the pill bugs, including their habitat, identifying features, and their scientific classification (Webber et al., 2013). This information was designed to bolster the

cover story (of preparing participants for their role of exterminator), as “real exterminators are very knowledgeable about the types of insects they work with” in order to “familiarize themselves” with the particular insects in the cups and increase their knowledge.

After reading the information packet, participants watched the video on the laptop (which demonstrated exactly how the task was supposed to be completed). In the video, a female researcher in a lab coat explained and demonstrated the extermination task, as well as provided an explanation of why the use of the grinder was necessary. The video specifically identified all of the items on the table, provided a step-by-step mimed demonstration of the extermination, and ended with instructions for participants. I utilized the demonstration video to circumvent the need for the experimenters to enter the “extermination area,” thereby reducing experimenter influence on participants’ killing decisions (and keeping the experimenters blind to conditions). At the end of the video, participants were informed they needed to dump each of the bugs into the tube on the extermination machine, one at a time, and they only needed to deposit as many insects as they thought necessary in order to “get a feel for what real exterminators experience.” This means participants chose the number of bugs they wished to kill, up to 10 (the number of bugs available in front of them).

Once all of the insects had been placed into the machine, participants were then instructed to press the (“activation”) button on the top of the machine for a minimum of three seconds to ensure the blades in the grinder moved fast enough to complete the extermination effectively (Webber et al., 2013). In reality, the tubing on the side of the machine runs directly against the outside of the grinder, rather than into the machine. When participants placed the bugs into the tube, they simply waited against the edge of the grinder, inside the tube. Thus, no bugs were harmed during this task, and participants are merely led to think they are

exterminating the insects. To add to the realism, the extermination machine contained pieces of Styrofoam so when participants pressed the activation button, they heard the blades in the grinder spinning and making noise, ostensibly grinding the bugs they had dumped. Once participants read the information packet and watched the demonstration video, they then performed the extermination task privately. Although the experimenter left the participant alone to complete the extermination task, they could hear the noise from the grinder through the door once participants pressed the activation button, and thus verified that they completed the task.

All participants then returned to the original room to complete the final packet of questionnaires “about what they did” during the extermination task, including the dependent measures of guilt and shame as indicators of psychological distress. The order of the materials for the final packet was: (1) the TRGI, (2) a brief filler scale assessing mood (PANAS), (3) the SSGS, and (4) the single item of willingness to kill. At this point, the researcher also informed participants that at the end of the packet there was information pertaining to a raffle they were eligible to enter as a reward for completing the study. Further instructions were provided in writing, but participants were given a roll of raffle tickets and told to fill out their name and email address on as many tickets as they wished, and to enter the raffle as many times as they liked. After participants had filled them out, they placed the raffle tickets into a separate box with a hole cut in the top so the researcher could not see how many tickets the participant had entered until the participant left. After each session, the experimenter emptied the box and counted the number of tickets for each participant. After explaining the tasks, the experimenter then left the room so the participant could complete the packet of dependent measures and fill out the raffle tickets in private. Upon completion of all tasks, the researcher probed participants

for suspicion (with particular attention to the extermination task, and whether they believed they had killed the bugs) and thoroughly debriefed them.

## Results

### Manipulation Checks

To assess whether the self-affirmation manipulation worked as intended, three independent raters read each essay and coded the responses on a number of dimensions. In a procedure adapted from past research (e.g., Williams, Schimel, Hayes, & Usta, 2014), raters used a 7-point scale (with higher numbers meaning greater agreement) to code the extent to which participants: (1) conveyed the value is important to them, (2) had their beliefs affirmed by writing about the value, (3) derived confidence as a consequence of writing the essay, and (4) wrote about themselves. I then computed Intraclass correlations (ICCs) across the three raters for each of the four questions. The ICCs all indicated acceptable levels of agreement (Fleiss, 1981) between raters for all four items (two-way mixed effects model, absolute agreement, all ICCs > .92). I then conducted one-way between-subjects ANOVAs on each of the four items to determine whether self-affirmation led to higher ratings compared to no affirmation, as predicted. The ANOVAs revealed participants in the affirmation condition indicated the value was more important to the participant,  $F(1, 46) = 110.19, p < .001; M = 6.06, SD = .55$ , versus  $M = 3.35, SD = 1.12$ . The affirmation essays were also rated as being more affirming to participants' beliefs,  $F(1, 46) = 245.95, p < .001; M = 5.68, SD = .66$ , versus  $M = 2.20, SD = .84$ , as well as led to increased confidence in participants' value as a person,  $F(1, 46) = 261.91, p < .001; M = 5.64, SD = .72$ , versus  $M = 2.11, SD = .77$ . Finally, the authors of the affirmation essays also referenced the self more frequently,  $F(1, 46) = 149.83, p < .001; M = 5.72, SD = .80$ , versus  $M = 2.15, SD = 1.16$ . These results indicate the self-affirmation manipulation was

effective in not only priming participants to think about a cherished attribute of the self, but also affirmed participants' values and beliefs.

### **Gender Effects**

I subjected all variables to condition x gender ANOVAs. Results revealed no significant effects for the number of bugs killed, the number of raffle tickets taken, the total score for the RSES, the total score of the TRGI or any of its subscales, the SSGS or any of its subscales, or the negative affect subscale of the PANAS (all  $ps > .05$ ). However, there was a main effect of gender for willingness to kill,  $F(1, 38) = 7.06, p = .01, partial \eta^2 = .16$ , where men indicated they were more willing to kill the bugs ( $M = 5.17, SD = 2.43$ ) than women ( $M = 3.13, SD = 2.59$ ). Additionally, there was a main effect of gender on the mean of the positive subscale of the PANAS,  $F(1, 39) = 7.67, p = .01, partial \eta^2 = .16$ , such that men reported higher positive affect ( $M = 2.34, SD = .70$ ) than women ( $M = 1.81, SD = .49$ ). I will further address these gender effects in the Discussion; however, because no other gender effects were found for any other variables, it is removed from the remaining analyses in the present study.

### **Effect of Condition on Supplemental Measures (Self-Esteem and Affect)**

To examine whether the self-affirmation manipulation influenced participants' feelings of self-worth or affect, I conducted one-way between-subjects ANOVAs on the mean scores for the RSES ( $\alpha = .92$ ), and for both the positive ( $\alpha = .86$ ) and negative ( $\alpha = .87$ ) subscales of the PANAS. Although I included these measures (self-esteem and affect) as filler items to bolster the cover story (and I therefore had no specific hypotheses regarding the outcome), I examined each of the three in turn in the interest of thoroughly exploring the data. Results revealed the main effect of condition on the RSES was not significant,  $F(1, 44) = .06, p = .80, partial \eta^2 = .00$ . Participants in the affirmation condition reported similar levels of self-esteem ( $M = 3.12, SD =$

.58) to those in the no affirmation condition ( $M = 3.08$ ,  $SD = .60$ ). There was also no effect of condition on positive affect,  $F(1, 44) = .07$ ,  $p = .79$ ,  $partial \eta^2 = .00$ , however, there was a significant main effect on negative affect,  $F(1, 44) = 11.02$ ,  $p < .01$ ,  $partial \eta^2 = .20$ . Participants in the affirmation condition reported slightly lower positive affect ( $M = 1.98$ ,  $SD = .70$ ) than those in the no affirmation condition ( $M = 2.03$ ,  $SD = .59$ ), although this difference was not significant. Those in the affirmation condition also reported significantly lower negative affect ( $M = 1.88$ ,  $SD = .63$ ) than those in the no affirmation condition ( $M = 2.57$ ,  $SD = .78$ ).

### **Effect of Condition on Self-Report Distress**

I calculated mean scores for each subscale of the TRGI (Distress, Guilt Cognitions, Hindsight, Wrongdoing, and Justification), and the SSGS (Shame, Guilt, and Pride) to examine whether engaging in self-affirmation (or not) buffers against the distress caused by completing a task that may violate one's morals. I then created a distress composite score to assess overall self-report guilt and distress after calculating standardized scores (i.e.,  $z$  scores) for each variable (i.e., the means of the subscales) and summing them together. I then subjected this self-report distress composite score ( $\alpha = .95$ ) to a one-way between-subjects ANOVA, which resulted in a significant main effect of condition,  $F(1, 41) = 7.13$ ,  $p = .01$ ,  $partial \eta^2 = .15$ . Contrary to the hypothesis, participants in the affirmation condition experienced significantly less distress ( $M = 3.05$ ,  $SD = 5.19$ ) than those in the no affirmation condition ( $M = 2.02$ ,  $SD = 7.15$ ).

I then analyzed each distress variable (including the totals for the TRGI and SSGS) separately to examine which aspects of distress may have been most influenced by the self-affirmation manipulation and extermination task (see Table 1 for means and statistics for all variables). One-way between-subjects ANOVAs revealed a significant effect of condition for 8 of the variables, including the total scale of the TRGI ( $\alpha = .94$ ), as well the Distress, Guilt

Cognitions, and Hindsight subscales, the total scale of the SSGS ( $\alpha = .94$ ), and all three subscales (Shame, Guilt, and Pride). The mean of the Justification subscale of the TRGI was marginally significant; however, the Wrongdoing subscale was the only non-significant finding.

Importantly, the pattern of results is the same for each variable (regardless of significance level) as the overall distress composite, such that self-affirmation led to less guilt and distress than no affirmation after killing (see Table 1).

Table 1

*Means, Standard Deviations, Cell Size, and Test Statistics for Each Individual Distress Variable from ANOVA on Self-Report Distress Composite, Including Total Scores for TRGI and SSGS (Study 1)*

Measure	Condition Means		Test Statistics
	Self-Affirmation	No Affirmation	ANOVA
TRGI Total	2.30 (.76) <i>n</i> = 23	2.95 (.99) <i>n</i> = 21	$F(1, 42) = 5.87, p = .02, \text{partial } \eta^2 = .12$
Distress	2.05 (1.22) <i>n</i> = 24	3.04 (1.51) <i>n</i> = 21	$F(1, 43) = 5.83, p = .02, \text{partial } \eta^2 = .12$
Guilt Cognitions	2.42 (.75) <i>n</i> = 23	3.00 (.91) <i>n</i> = 22	$F(1, 43) = 5.59, p = .02, \text{partial } \eta^2 = .12$
Hindsight	2.78 (.80) <i>n</i> = 24	3.33 (.97) <i>n</i> = 22	$F(1, 44) = 4.50, p = .04, \text{partial } \eta^2 = .09$
Wrongdoing	1.96 (.93) <i>n</i> = 24	2.22 (.89) <i>n</i> = 22	$F(1, 44) = .94, p = .34, \text{partial } \eta^2 = .02$
Justification	2.67 (1.19) <i>n</i> = 23	3.36 (1.24) <i>n</i> = 22	$F(1, 43) = 3.64, p = .06, \text{partial } \eta^2 = .08$
SSGS Total	2.21 (.70) <i>n</i> = 23	2.87 (.99) <i>n</i> = 22	$F(1, 43) = 6.69, p = .01, \text{partial } \eta^2 = .14$
Shame	1.54 (.67) <i>n</i> = 24	2.07 (.92) <i>n</i> = 22	$F(1, 44) = 5.04, p = .03, \text{partial } \eta^2 = .10$
Guilt	1.90 (.89) <i>n</i> = 23	2.73 (1.34) <i>n</i> = 22	$F(1, 43) = 5.95, p = .02, \text{partial } \eta^2 = .12$
Pride	3.23 (.91) <i>n</i> = 24	3.82 (.95) <i>n</i> = 22	$F(1, 44) = 4.69, p = .04, \text{partial } \eta^2 = .10$

*Note.* Standard deviations appear in parentheses.

### **Effect of Condition on Willingness to Kill and Number Killed**

Although the distress measures were the main dependent variables of interest, I also examined the single continuous item of willingness to kill and the number of insects killed using one-way between-subjects ANOVAs to determine whether the self-affirmation manipulation affected these indices. There was no effect of condition on willingness to kill,  $F(1, 43) = .28, p = .60, \text{partial } \eta^2 = .01$ , although participants in the affirmation condition reported they were slightly more willing to kill the bugs ( $M = 4.08, SD = 2.62$ ) than those in the no affirmation condition ( $M = 3.67, SD = 2.71$ ). The effect of condition on the number of bugs killed was also non-significant,  $F(1, 44) = .56, p = .46, \text{partial } \eta^2 = .01$ , with participants in the affirmation condition killing slightly more bugs ( $M = 2.88, SD = 3.29$ ) than those in the no affirmation condition ( $M = 2.27, SD = 1.96$ ). It is also interesting to note that although participants in the affirmation condition reported being slightly more willing to kill, more participants in this condition chose to kill zero bugs ( $n = 10$ ), compared to the no affirmation condition ( $n = 6$ ), however some of these participants were excluded from analyses for other reasons (e.g., suspicion).

### **Effect of Condition on Behavioral Guilt (Self-Punishment)**

I also assessed the number of raffle tickets participants took as a means of examining self-punishment (as a proxy for behavioral guilt and distress). I subjected the number of raffle tickets to a one-way between-subjects ANOVA, which was non-significant,  $F(1, 43) = 2.40, p = .13, \text{partial } \eta^2 = .05$ . Participants in the affirmation condition took slightly fewer raffle tickets ( $M = 2.35, SD = 2.31$ ) than those in the no affirmation condition ( $M = 4.14, SD = 5.02$ ).

## **Discussion**

Contrary to the hypothesis, the findings from Study 1 showed that engaging in self-affirmation buffered against the guilt and distress typically observed after completing a task that

violates one's moral standards (i.e., killing). Across a range of self-report measures, after self-affirming, participants experienced less guilt and distress in response to the extermination task. More specifically, participants in the self-affirmation condition reported lower levels of trauma-related guilt, as well as shame and guilt, as assessed by two separate measures and an overall self-report distress composite score.

Additionally, self-affirmed participants reported lower negative affect than did non-affirmed participants. This last finding for negative affect is in line with the results obtained on the various distress indices, suggesting self-affirmation led to lower distress than no affirmation. These results are consistent with previous research showing self-affirmation can provide an alternative source of self-validation (or can restore one's feeling of integrity or "goodness") and thereby serve to buffer against feelings of guilt and shame resulting from completing a task that goes against one's moral code.

One unexpected finding was the effect of gender, which revealed men were more willing to kill, and experienced higher positive affect than women did. These findings were only unexpected in that I was not predicting any gender effects; however, neither finding is entirely surprising in and of itself. Previous research (e.g., Webber, 2015) using the same bug-killing paradigm reported similar effects with gender, such that men were able to justify killing more so than women, and were more likely to "opt-in" to complete the extermination task when given the choice. This effect of men opting-in more than women is consistent with the finding in the present study of men reporting being more willing to kill. Moreover, these effects of gender align with a great deal of past research (see Archer, 2004 or Card, Stucky, Sawalani, & Little, 2008 for meta-analyses), showing men use more overt forms of aggression than women do. If men are generally more overtly aggressive than women are, it makes sense that they would also

be more willing to kill the bugs in an admittedly aggressive manner (i.e., using the grinder).

Along the same lines, if one is more willing to complete the task, it follows that it may have had less of an effect on one's negative mood. This finding does not necessarily mean that men were completely unfazed by the task, as even their "higher" level of positive affect was still below the midpoint (at 2.34 on a 5-point scale) and no gender effects were obtained on any of the main indices of guilt or distress, suggesting men still found the task troubling.

Two other findings (or lack thereof) are also worth noting. Neither the number of bugs participants chose to kill nor the number of raffle tickets taken differed between conditions. However, as noted, both of these indicators were supplemental to the main dependent measures and I did not have specific predictions regarding how self-affirmation would influence the number of bugs killed in particular. Moreover, although non-significant, the pattern of means shows self-affirmation affected the number of bugs killed in a similar fashion as the other dependent measures. Participants in the self-affirmation condition killed more bugs than those in the no affirmation condition, yet reported less distress. Unexpectedly, however, self-affirmed participants took fewer raffle tickets than non-affirmed participants did. Although this finding is contrary to my hypothesis, the pattern was non-significant, and many participants expressed suspicion regarding this task (which I will discuss later in the paper). Although it would have been ideal to obtain findings on all measures that were included, the main dependent variables (guilt and distress) were significant, and the effect of self-affirmation on guilt and distress after killing was therefore deemed worthy of further exploration.

As noted, the purpose of Study 1 was to assess the hypothesis that self-affirmation would increase the guilt and distress of killing by highlighting the discrepancy between positive self-attitudes and immoral behavior. Study 1 instead revealed self-affirmation had the opposite effect

of buffering the distress of killing, an effect that is more in line with prior self-affirmation research. However, some limitations make the results of Study 1 difficult to interpret. First, and importantly, there were issues with participant compliance regarding the bug-killing task. As may be expected, some participants did not want to complete the extermination task and revealed to the researcher during the debriefing that they continued and decided to “go through” with the study only because they were allowed to choose how many bugs they needed to kill. Because participants could kill as many bugs as they needed to feel like exterminators, many participants chose to kill zero bugs. These individuals still technically “participated” within the constraints presented to them, but this manner of participation likely placed them into a very different psychological state than participants who chose not to kill.

Clearly, if participants are not actually killing any bugs, they are not likely to experience distress because of their actions. In fact, they may take pride in their decision to do the “right thing” by not killing the bugs. Moreover, giving participants a choice in how many bugs to kill may have allowed them to regulate their degree of dissonance to a tolerable level. It seems reasonable that many participants chose to kill just enough bugs (1 or 2) to make them feel like good research participants, but not enough to arouse distress over their actions. Because the main dependent variables examined guilt and distress specifically in response to killing, this methodological issue likely rendered Study 1 a poor test of the hypothesis. It is difficult, therefore, to draw firm conclusions about how self-affirmation affects the distress of killing based on the results of Study 1. To address this methodological issue and provide a better test of the key hypothesis, I conducted a second study that required participants to kill a specified number of bugs, rather than allowing them to choose how many they wished to kill.

A second limitation of Study 1 was that it utilized a behavioral measure of distress that did not work as well as intended. As discussed, participants were told they could take as many tickets to enter into a raffle as they desired, as a “reward” for completing the bug-killing task (which presumably many of them did not want to do). Although this task had been used in previous research (e.g., Webber et al, 2013), it may not be a reliable measure (e.g., these researchers had to transform the raffle ticket scores due to problems they encountered). One potential explanation for the lack of any effect on this measure was many participants were highly suspicious of the task (e.g., they did not believe they could really take as many tickets as they desired). Based on their post-experimental comments, many participants took few or no tickets to avoid looking as though they had been duped, which led to a floor effect. To address this limitation, Study 2 employed a different behavioral measure of guilt that was more subtle and less likely to arouse suspicion.

## **Study 2**

Although the results of Study 1 were inconclusive given the aforementioned methodological issues, the fact that many participants opted not to kill or killed very few bugs when given the choice suggests they found the prospect of killing bugs to be highly distressing. In Study 2, I therefore retained the bug extermination task, but modified the procedure such that participants were required to kill exactly ten bugs. Participants in Study 2 began by completing the same self-affirmation (vs. control) task as in Study 1, followed by the modified extermination task, and the self-report measures of guilt and shame. However, instead of having participants choose raffle tickets as a measure of guilt (i.e., self-punishment), I assessed hand washing as a subtle behavioral measure of guilt. This task has been used in previous research to examine both obsessive-compulsive type guilt as well as guilt over behaving immorally (e.g., Strachan et al.,

2007), demonstrating that participants who feel more guilt over something they have done spend more time washing their hands than those who do not feel guilty (Cougles, Goetz, Hawkins, & Fitch, 2012). This research suggests that compulsively washing one's hands serves to "cleanse" oneself physically of the immoral behavior, thereby "washing away" psychological guilt and distress.

I therefore used a hand-washing task as a discreet indicator of how guilty and distressed participants might have been feeling after the extermination task. The hope was this behavioral measure would align with the self-report assessments, thereby providing converging evidence for the effect of self-affirmation on the distress caused by killing. Moreover, the self-report measures used in the present studies (the TRGI and SSGS) are fairly face valid, meaning they look like what they are assessing. It is easy for participants to deduce that we are interested in examining how guilty they are feeling about the bug-killing task because the questions are straightforward (and indeed participants often commented on this fact during debriefing). As such, I used the subtle hand washing measure to reduce socially desirable responding, suspicion, and/or demand characteristics. After completing the bug extermination task, the experimenter suggested to the participants that they might want to wash their hands after handling the bugs. The experimenter then surreptitiously recorded the amount of time, soap, and paper towels participants used to wash their hands. I predicted that self-affirmed participants would report more guilt and shame after killing bugs and engage in more hand washing behavior.

## **Method**

### **Participants**

Sixty-six undergraduates (18 men, 43 women, 5 unreported) from the University of Alberta were recruited in exchange for partial credit towards the completion of an introductory

psychology class. Participants were at least 18 years of age ( $M = 18.90$ ,  $SD = 2.03$ ). However, seven participants were excluded from analyses due to suspicion regarding the tasks (e.g., they suspected they were not actually killing the bugs) and/or research hypotheses, not following instructions, or experimenter error. A final sample of 59 (15 men, 39 women, 5 unreported) participants were included in all analyses for which they had complete data ( $M_{age} = 18.98$ ,  $SD_{age} = 2.12$ ). All requirements established by the Ethics Review Board at the University of Alberta were satisfied, as well as the APA ethical standards for participation with human research subjects.

### **Research Design**

Study 2 used the same one-way between-subjects design as Study 1, manipulating one independent variable with two conditions. Participants were randomly assigned to either the self-affirmation condition ( $n = 29$ ) or no affirmation control ( $n = 30$ ), and the dependent variables included several self-report measures of distress (e.g., guilt and shame) as well as a behavioral measure of guilt. The present study also employed a double-blind procedure in which neither the experimenter nor the participants were aware of the experimental condition to which participants had been assigned.

### **Materials and Procedure**

The materials and procedure for Study 2 were largely the same as those employed in Study 1. All participants completed the study individually. Upon arrival, the researcher greeted participants in the hallway and brought them into the laboratory where they completed an informed consent form after a brief introduction to the study. As with Study 1, the researcher informed participants they would complete a study investigating how exterminators deal with insects, and that we were interested in personality and exterminators. The researcher told

participants the study involved a bug extermination task as well as answering some questions regarding their experience. Participants then completed a series of tasks, independently and at their own pace.

All participants first completed a packet of “personality measures” containing the experimental manipulation (i.e., self-affirmation or no affirmation control writing task) to which they had been randomly assigned. Next, participants were ushered by the researcher from the main room to the “extermination area” in an adjacent room to complete the extermination task. I utilized the same set-up and instructions from Study 1 here. Specifically, the “extermination area” contained a desk and all the materials required to complete the task. The experimenter instructed the participants to: (1) read through the information packet, (2) watch the video on the laptop, (3) complete the extermination task as demonstrated in the video, and (4) return to the main room for further instructions. However, instead of allowing participants to choose the number of bugs to kill (Study 1), I modified the instructions, such that the video instructed participants to kill all 10 bugs. Upon completing the extermination task, all participants then returned to the main lab.

As discussed, another important modification in Study 2 was the inclusion of a different behavioral measure of guilt using a hand-washing task. For this task, once participants had returned to the main room, the researcher provided them with an opportunity to wash their hands before they filled out the final dependent questionnaires. The experimenter met the participant at the door to the main room when they returned after the extermination task and commented that they might want to “get cleaned up” after handling the bugs. The experimenter indicated a sink in the room where there were paper towels and soap set up for participants. As the participants washed up, the experimenter left the room “to get the next packet of questionnaires” and secretly

timed how long the participants spent washing their hands in seconds, beginning once participants turned the tap on and stopping when they turned the tap off. Once the experimental session was over, the researcher weighed how much soap each participant used. I then subtracted the amount at the end of the session (post-weight) from the amount in the bottle before the session began (pre-weight). The researcher also counted the number of paper towels each participant used at the end of each session (there were always three paper towels in the trash before each participant arrived so they would notice other participants had also washed their hands).

Once participants finished washing their hands (the researcher listened at the door with the timer and waited a few seconds after the water stopped), the researcher opened the door and entered with the final packet of questionnaires, including the dependent measures of guilt and shame as indicators of psychological distress. The order of the materials for the final packet was: (1) the TRGI, (2) a scale assessing mood (PANAS), and (3) the SSGS. The final item used in Study 1 asking participants about how willing they were to complete the extermination task was omitted from the remaining studies as it was not deemed particularly useful or relevant now that all participants were instructed to kill the same number of bugs. After explaining the final packet, the researcher then left the room so the participant could complete the questionnaires in private. Upon finishing all tasks, the researcher probed participants for suspicion and thoroughly debriefed them.

## **Results**

### **Manipulation Checks**

As with the previous study, three independent raters read each essay and coded the responses on the same four questions used in Study 1 to assess whether the self-affirmation

manipulation was effective. I then computed Intraclass correlations (ICCs) across the three raters for each of the four questions. The ICCs all indicated good levels of agreement (Fleiss, 1981) between raters for all four items (two-way mixed effects model, absolute agreement, all ICCs > .86). I then conducted one-way between-subjects ANOVAs on each of the four items to determine whether self-affirmation led to higher ratings, as predicted. The ANOVAs revealed participants in the affirmation condition indicated the value was more important to the participant,  $F(1, 59) = 299.35, p < .001; M = 6.44, SD = .61$ , versus  $M = 3.07, SD = .86$ . The affirmation essays were also rated as being more affirming to participants' beliefs,  $F(1, 59) = 160.86, p < .001; M = 6.09, SD = .87$ , versus  $M = 3.32, SD = .81$ , as well as increased participants' confidence in their value as a person,  $F(1, 59) = 164.07, p < .001; M = 5.78, SD = .96$ , versus  $M = 2.90, SD = .76$ . Finally, the authors of the affirmation essays referenced the self more frequently,  $F(1, 59) = 92.81, p < .001; M = 5.64, SD = 1.68$ , versus  $M = 1.81, SD = 1.37$ . Taken together, these results indicate that the self-affirmation manipulation was effective not only in priming participants to think about a cherished attribute of the self, but also served to affirm participants' beliefs and values as a person.

### **Gender Effects**

I subjected all variables to condition x gender ANOVAs and a significant condition x gender interaction was found for the self-report distress composite score,  $F(1, 44) = 7.69, p = .01, partial \eta^2 = .15$ , as well as for many of the individual variables (e.g., the total score of the TRGI). In general, the pattern was such that men reported less distress in the affirmation condition than women did, but the pattern was reversed in the no affirmation condition (with men reporting higher distress than women). However, the gender effect was not present for the behavioral distress composite score,  $F(1, 46) = 1.46, p = .23, partial \eta^2 = .03$ , nor was there a

main effect of gender,  $F(1, 46) = .00, p = .98, \text{partial } \eta^2 = .00$ . Because the self-report measures revealed this pattern, but the behavioral measure did not, I dropped gender from all remaining analyses. I will further address this gender effect in the Discussion.

### **Effect of Condition on Supplemental Measures (Self-Esteem and Affect)**

To examine whether the self-affirmation manipulation influenced participants' feelings of self-worth or affect, I conducted one-way between-subjects ANOVAs on the mean scores for the RSES ( $\alpha = .89$ ), and for both the positive ( $\alpha = .85$ ) and negative ( $\alpha = .85$ ) subscales of the PANAS. As with Study 1, these measures (of affect and self-esteem) were merely included as filler items to maintain the cover story, and I therefore did not have specific hypotheses concerning them, especially given that there was no effect on self-esteem in Study 1. However, in the interest of thoroughly exploring all facets of the data, I again examined each of the three items. Results of the analyses revealed there was a marginally significant effect of condition on self-esteem,  $F(1, 55) = 3.75, p = .06, \text{partial } \eta^2 = .06$ , such that participants in the affirmation condition reported slightly lower self-esteem ( $M = 3.00, SD = .54$ ) than those in the no affirmation condition ( $M = 3.27, SD = .49$ ).

In line with what was observed in Study 1, there was no effect of condition on positive affect,  $F(1, 56) = .29, p = .59, \text{partial } \eta^2 = .01$ , however, there was an effect on negative affect,  $F(1, 55) = 5.58, p = .02, \text{partial } \eta^2 = .09$ . Participants in the affirmation condition reported slightly lower positive affect ( $M = 2.17, SD = .78$ ) than those in the no affirmation condition ( $M = 2.27, SD = .66$ ), although this difference was not significant. Participants in the affirmation condition also reported significantly higher negative affect ( $M = 2.31, SD = .77$ ) than those in the no affirmation condition ( $M = 1.86, SD = .69$ ), in contrast to what was observed in Study 1.

### **Effect of Condition on Self-Report Distress**

I created a distress composite score ( $\alpha = .94$ ) in the same way as the previous study to assess participants' overall guilt and distress on the self-report measures. I then computed a one-way between-subjects ANOVA on this self-report distress score to examine whether engaging in self-affirmation (or not) affects distress, which revealed a significant main effect of condition,  $F(1, 49) = 5.23, p = .03, \text{partial } \eta^2 = .10$ . Participants in the affirmation condition experienced significantly more distress ( $M = 2.24, SD = 6.88$ ) than those in the no affirmation condition ( $M = -1.92, SD = 6.09$ ), in line with the hypothesis.

I then analyzed each individual distress variable separately to examine which aspects of distress may have been most influenced by the self-affirmation manipulation and extermination task. See Table 2 for means and statistics for all variables. I subjected all variables to a one-way between-subjects ANOVA, and found significant differences among 5 of the variables, including the means of the Distress, and Justification subscales of the TRGI, and the mean of the total scale of the SSGS ( $\alpha = .93$ ) and the means of two of the three subscales (Shame and Guilt). The mean of the total scale of the TRGI ( $\alpha = .94$ ) as well as the mean of the Guilt Cognitions subscale were marginally significant, however, the means of the Hindsight and Wrongdoing subscales of the TRGI and the Pride subscale of the SSGS did not reach significance. Importantly, the pattern of results was the same for each variable (regardless of significance level) as the overall distress composite. In contrast to the results of Study 1, self-affirmation led to more guilt and distress than no affirmation after killing (see Table 2).

Table 2

*Means, Standard Deviations, Cell Size, and Test Statistics for Each Individual Distress Variable from ANOVA on Self-Report Distress Composite, Including Total Scores for TRGI and SSGS (Study 2)*

Measure	Condition Means		Test Statistics
	Self-Affirmation	No Affirmation	ANOVA
TRGI Total	2.82 (.80) <i>n</i> = 27	2.43 (.83) <i>n</i> = 28	$F(1, 53) = 3.23, p = .08, \text{partial } \eta^2 = .06$
Distress	2.82 (1.25) <i>n</i> = 29	2.16 (1.13) <i>n</i> = 29	$F(1, 56) = 4.49, p = .04, \text{partial } \eta^2 = .07$
Guilt Cognitions	2.87 (.73) <i>n</i> = 27	2.51 (.81) <i>n</i> = 28	$F(1, 53) = 3.06, p = .09, \text{partial } \eta^2 = .06$
Hindsight	2.95 (.84) <i>n</i> = 29	2.74 (.80) <i>n</i> = 30	$F(1, 57) = 1.01, p = .32, \text{partial } \eta^2 = .02$
Wrongdoing	2.16 (.83) <i>n</i> = 28	1.96 (.77) <i>n</i> = 29	$F(1, 55) = .94, p = .34, \text{partial } \eta^2 = .02$
Justification	3.55 (.86) <i>n</i> = 28	3.01 (1.14) <i>n</i> = 29	$F(1, 55) = 4.14, p = .05, \text{partial } \eta^2 = .07$
SSGS Total	2.90 (.90) <i>n</i> = 25	2.35 (.65) <i>n</i> = 30	$F(1, 53) = 6.88, p = .01, \text{partial } \eta^2 = .12$
Shame	2.04 (.79) <i>n</i> = 27	1.55 (.72) <i>n</i> = 30	$F(1, 55) = 5.84, p = .02, \text{partial } \eta^2 = .10$
Guilt	2.87 (1.19) <i>n</i> = 27	1.92 (.89) <i>n</i> = 30	$F(1, 55) = 11.65, p = .001, \text{partial } \eta^2 = .18$
Pride	3.81 (.89) <i>n</i> = 27	3.57 (.80) <i>n</i> = 30	$F(1, 55) = 1.17, p = .29, \text{partial } \eta^2 = .02$

*Note.* Standard deviations appear in parentheses.

### **Effect of Condition on Behavioral Guilt (Hand Washing)**

In line with creating a self-report distress composite, I also formed a behavioral guilt composite of hand washing by combining the amount of time spent, the amount of soap used (subtracting the post-weight from the pre-weight), and the number of paper towels used. After standardizing each of these three components into *z* scores, I summed them to form a composite measure of hand washing. I then subjected this hand washing composite score ( $\alpha = .58$ ) to a one-way between-subjects ANOVA to examine the effect of condition on overall behavioral guilt. Results revealed a significant main effect of condition,  $F(1, 53) = 7.26, p = .01, \text{partial } \eta^2 = .12$ .

Consistent with what was observed on the self-report measures, participants in the affirmation condition ( $M = .82$ ,  $SD = 2.48$ ) displayed higher levels of guilt than those in the no affirmation condition ( $M = -.72$ ,  $SD = 1.74$ ).

I then analyzed each variable in the hand washing composite separately to examine which aspects of behavioral distress may have been most influenced by the self-affirmation manipulation and extermination task. Means and statistics for all variables, including both the raw (mean amount of time, mean amount of soap, and mean number of paper towels) as well as the transformed  $z$  scores for each are presented in Table 3. I subjected each variable to a one-way between-subjects ANOVA, and found a significant effect for soap used (in grams), and found marginal effects for both the amount of time spent (in seconds) and the number of paper towels used. In each case, the pattern of results suggests participants in the self-affirmation condition displayed more distress than those in the control condition after killing the bugs.

Table 3

*Means, Standard Deviations, Cell Size, and Test Statistics for Each Individual Distress Variable from ANOVA on Behavioral Distress Composite, Including Both Raw and Standardized Z Scores for Time, Towels, and Soap (Study 2)*

Measure	Condition Means		Test Statistics
	Self-Affirmation	No Affirmation	ANOVA
Time	20.96 (10.49) $n = 26$	16.77 (7.39) $n = 30$	$F(1, 54) = 3.05$ , $p = .09$ , $partial \eta^2 = .05$
Towels	1.79 (.63) $n = 28$	1.48 (.63) $n = 29$	$F(1, 55) = 3.28$ , $p = .08$ , $partial \eta^2 = .06$
Soap	.96 (.69) $n = 28$	.65 (.31) $n = 30$	$F(1, 56) = 5.07$ , $p = .03$ , $partial \eta^2 = .08$
Time $z$ Scores	.28 (1.17)	-.18 (.82)	
Towels $z$ Scores	.21 (.96)	-.25 (.96)	
Soap $z$ Scores	.29 (1.28)	-.28 (.57)	

*Note.* Standard deviations appear in parentheses.

## Discussion

Consistent with the hypothesis, Study 2 demonstrated that self-affirmation increased participants' guilt and distress caused by killing. In fact, nearly every index of guilt and distress that was assessed demonstrated this pattern, such that self-affirmed participants reported and displayed higher levels of guilt and distress after the extermination task, compared to those in the control condition. Interestingly, the more discreet behavioral measure of guilt also confirmed this pattern, such that self-affirmed participants spent more time, used more soap, and used more paper towels when washing their hands after killing the bugs than participants who had not self-affirmed.

It appears that altering the bug-killing paradigm so all participants killed a specified number of bugs does indeed lead to more guilt and distress, as expected. In the present study then, participants are no longer able to justify their actions because killing 10 bugs is simply too many to preclude dissonance from being aroused by choosing to kill few or no bugs (as in Study 1), and they perhaps felt they had "crossed a line" for what they perceived as acceptable behavior. Thus, in Study 1, when participants are allowed to choose the number of bugs to kill, engaging in self-affirmation and restoring one's sense of self-integrity appeared to help buffer against the distress that may be experienced because of engaging in an act that would normally violate one's moral standards. However, as noted, caution is urged in interpreting these findings suggesting self-affirmation may reduce distress under these conditions, as the methodological issues encountered in Study 1 make it difficult to draw firm conclusions (e.g., participants may not have actually been distressed if they did not actually kill any bugs). In the present study, however, when I addressed the methodological limitations and removed the "choice" to kill, self-affirmation causes an increase in guilt and distress, as predicted.

Similar to Study 1, there was no effect on self-esteem or positive affect, but there was an effect of condition on negative affect. Again, although I did not have specific predictions regarding these ancillary measures, the finding on negative affect is not surprising and fits well with the remainder of the results. Participants in the self-affirmation condition reported greater negative mood than control participants. As many of the items on the negative subscale of the PANAS refer to feeling anxious, guilty, and upset, this finding aligns with what might be expected.

Other notable findings include the gender effect mentioned at the beginning of the results section, in which there was an interaction between gender and condition on many of the self-report measures of distress. Overall, the general pattern is women in the self-affirmation condition reported being more distressed than in the no affirmation condition, whereas men reported being less distressed in the self-affirmation condition. Importantly, however, there is no effect of gender on the behavioral guilt measure of hand washing (i.e., both men and women are equally more distressed in the self-affirmation condition than the control condition). On the surface, this pattern of findings may seem puzzling. However, after careful consideration, there is a plausible explanation for these results.

First, and most importantly, given the diverging pattern between the self-report and behavioral measures, one possible interpretation is men were simply less willing to “report” that they were upset or distressed by the extermination task. Perhaps they wished to “put on a brave face” or appear more masculine (i.e., “macho”) in front of the researcher (all of whom were female). Given that the hand washing measure is much more inconspicuous and implicit (and indeed the researcher did not even introduce it as part of the study), this measure is perhaps more accurate and likely to tap participants’ true levels of distress. Speaking to the discreet nature of

this task, it is also worth noting that no participants indicated being suspicious about washing their hands as part of the study, much less as a measure of how guilty and distressed they felt. The fact that men and women behaved in a similar fashion on this task is telling. The gender effect therefore seems to be largely due to men engaging in self-presentation, wanting to present a macho façade, rather than to actual differences in levels of distress between the genders, as initially revealed on the self-report measures.

Another reason to question the findings with gender is the cell sizes between men and women across conditions are far from equal. Because there were only 15 men included in the analyses (compared to 39 women), and not every participant completed all measures, there were some analyses that only included full responses from four men in the self-affirmation condition (e.g., the self-report distress composite). This suggests comparisons between men and women were woefully underpowered, and basing conclusions about gender differences on these low and possibly unreliable cells sizes would seem ill advised. I therefore suggest caution be used in interpreting the gender effects. Of course, given that there was a consistent effect of gender observed on the self-report measures, I examined gender in each of the subsequent studies as well.

One more potential issue worth noting is the alpha level for the behavioral distress composite is lower than would normally be desired (and much lower than the self-report scales). Although a low alpha is not ideal, it is important to keep in mind that there were only three measures included in the distress composite (time, soap, and paper towels), which likely accounts for this finding. Indeed, the more items that are included in a reliability analysis, the higher the alpha level will generally be (Pallant, 2007). In addition, the only component of the hand-washing task that was significant on its own was the amount of soap participants used to

wash their hands. The other two (the amount of time spent washing and the number of paper towels used) were both marginal. However, two things are worth bearing in mind here. First, the pattern of means remains the same for each of the three components, such that self-affirmed participants behaved in a manner suggesting they were more distressed, by washing their hands for longer, using more soap, and using more paper towels than non-affirmed participants were. Thus, it can be argued that it would be more problematic if each measure was not only non-significant, but if the patterns also differed, which is not the case. Moreover, the fact that the individual components are not significant, yet the overall composite is, merely speaks to the power of aggregating similar items together and indeed is the reason a composite was created (in line with previous research; e.g., Strachan et al., 2007). Each individual behavioral measure may be somewhat unreliable, but combining all of these measures increases our power and therefore our ability to find meaningful information from the data (Warner, 2008).

Taken together, the results from Study 2 support the hypothesis that self-affirmation may backfire, leading to more guilt and distress after killing. However, given the contradictory findings between Studies 1 and 2, it seemed pertinent to conduct a third study to attempt to reconcile the diverging results and replicate the findings of Study 2, as a means of ensuring it was actually the difference in designs that accounts for the discrepancies. Additionally, the designs of Studies 1 and 2 were very simple, employing the use of only two conditions (self-affirmation or no affirmation control). It is important, therefore, to assess how the kill/no affirmation condition (employed in Studies 1 and 2) may differ from a no kill control condition. Because all participants in Study 2 killed bugs, it is unclear whether the kill/affirmation condition increased guilt and distress, or whether the kill/no affirmation condition decreased

guilt and distress through typical dissonance reduction efforts. As such, I expanded the design in Study 3 and incorporated a third, no kill control condition.

### **Study 3**

Study 3 was designed to build on the findings of Studies 1 and 2 and determine whether the contrasting pattern of results was indeed due to self-affirmation leading to increased guilt and distress upon having killed a pre-determined number of bugs. As such, Study 3 used the same method as Study 2 in which participants killed a standardized number of (10) bugs. Moreover, a no kill control condition was introduced to assess the differential effect of killing (vs. not killing) on guilt and distress. One final change was participants completed the self-affirmation manipulation after the extermination task, rather than before, to examine whether the timing of the manipulation has an effect on participants' distress.

I expected that Study 3 would replicate the findings of Study 2, such that participants who killed and engaged in self-affirmation (kill/affirmation condition) would experience more guilt and distress than participants in the kill/no affirmation and no kill/no affirmation conditions. Additionally, I expected that participants in the no kill/no affirmation condition would experience the lowest levels of distress compared to the other conditions, since these individuals were not in fact killing any bugs. As means of assessing these hypotheses, participants once again completed the extermination task, followed by the self-affirmation manipulation. Both self-report and hand washing measures assessed distress.

### **Method**

#### **Participants**

Ninety-two undergraduates (33 men, 58 women, 1 unreported) from the University of Alberta were recruited in exchange for partial credit towards the completion of an introductory

psychology class. Participants were at least 18 years of age ( $M = 19.03$ ,  $SD = 1.51$ ).

Unfortunately, however, 17 participants were excluded from analyses due to suspicion regarding the tasks (e.g., did not believe they were actually killing the bugs) and/or research hypotheses, not following instructions, or experimenter error. The bulk of these participants were excluded from the no kill control condition because they did not understand that they were not killing (or harming) the bugs, despite the instructions provided. I will address this limitation further in the discussion. A final sample of 75 (26 men, 48 women, 1 unreported) participants were included in all analyses for which they had complete data ( $M_{age} = 19.01$ ,  $SD_{age} = 1.52$ ). All requirements established by the Ethics Review Board at the University of Alberta were satisfied, as well as the APA ethical standards for participation with human research subjects.

### **Research Design**

I employed a similar one-way between-subjects design as in Studies 1 and 2, manipulating one independent variable. However, for the current study, there were three conditions, rather than two, as an additional control condition was introduced in which participants knew they were not actually killing the bugs. The first two conditions were the same as in Studies 1 and 2, in which participants killed and then completed the self-affirmation or control manipulation (kill/affirmation vs. kill/no affirmation), and the third condition was a no kill control condition, in which participants also completed the control manipulation (no kill/no affirmation). Participants were randomly assigned to either the kill/affirmation condition ( $n = 26$ ), kill/no affirmation ( $n = 25$ ), or no kill/no affirmation control ( $n = 24$ ), and the dependent variables included several self-report measures of distress (e.g., guilt and shame) as well as a behavioral measure of distress. The present study also employed a double-blind procedure in

which neither the experimenter nor the participants were aware of the experimental condition to which the participants had been assigned.

### **Materials and Procedure**

The materials and procedure for Study 3 were largely the same as those of Study 2. All participants completed the study individually. Upon arrival, the researcher greeted participants in the hallway and brought them into the laboratory where they completed an informed consent form after a brief introduction to the study. As with the other studies, the researcher told participants they would be completing a study examining how exterminators deal with insects, and that for this specific study we were interested in personality and exterminators. As such, the researcher informed participants the study involved completing a bug extermination task as well as answering some questions regarding their experience. Participants then completed a series of tasks independently, at their own pace.

Unlike in Studies 1 and 2, all participants completed the extermination task first, followed by the self-affirmation manipulation. After completing the consent form in the main room, the experimenter led participants to the adjacent room to the extermination area to complete the bug-killing task. I employed the same set-up as the previous studies here, although I modified the instructions slightly. Specifically, the “extermination area” contained a desk and all the materials required to complete the task. However, for this study, the personality measures (containing the experimental manipulation – self-affirmation or no affirmation control – to which they had been randomly assigned) were also sealed in an envelope on the desk. Participants were instructed to: (1) read through the information packet, (2) watch the video on the laptop, (3) complete the extermination task as demonstrated in the video, (4) open the packet and complete the personality measures, and then (5) return (with the envelope of personality measures) to the

main room for further instructions. The purpose of having participants complete the “personality packet” in the cubicle was to streamline the procedure and keep the final tasks the same as in Study 2 (i.e., having participants wash their hands immediately when they returned to the main room followed by the final packet of measures).

Thus, the primary difference between the present study and Studies 1 and 2 was participants completed the self-affirmation manipulation after the extermination task, rather than before. I also altered the video for the no kill condition so participants learned at the end (after the demonstration) that they would not actually be killing any bugs. These instructions appeared at the end of the demonstration explaining in writing that there was a barrier between the tubing and the grinder so no bugs would actually enter the machine. Everything else (for participants in the two kill conditions – affirmation and no affirmation) remained the same as in Study 2. All participants were instructed to kill (or, for the no kill condition, place into the grinder) 10 bugs and press and hold the activation button for at least three seconds.

As with the previous study, the researcher provided participants with an opportunity to wash their hands after they returned from the extermination area, as a measure of behavioral guilt. Once participants had finished washing their hands, the researcher opened the door and entered the main room with the final packet of questionnaires regarding what “they did” during the extermination task, containing, in order, the TRGI, PANAS, and SSGS. After explaining the final packet, the researcher then left the room so the participant could complete the final questionnaires in private. Upon completion of all tasks, the researcher probed participants for suspicion and thoroughly debriefed them.

## Results

### Manipulation Checks

To assess whether the self-affirmation manipulation worked as intended, three independent raters read each essay and coded the responses on the same four questions used in the previous studies. I then computed Intraclass correlations (ICCs) across the three raters for each of the four questions. The ICCs all indicated good levels of agreement (Fleiss, 1981) between raters for all four items (two-way mixed effects model, absolute agreement, all ICCs > .84). One-way ANOVAs conducted on each of the four items determined whether self-affirmation led to higher ratings, as predicted. The ANOVAs revealed participants in the affirmation condition indicated the value was more important to the participant,  $F(1, 75) = 1052.38, p < .001; M = 6.60, SD = .34$ , compared to those in the control condition,  $M = 2.41, SD = .61$ . The affirmation essays were also rated as being more affirming to participants' beliefs,  $F(1, 75) = 547.81, p < .001; M = 6.23, SD = .54$ , versus  $M = 2.84, SD = .62$ , as well as increasing participants' confidence in their value as a person,  $F(1, 75) = 645.29, p < .001; M = 5.92, SD = .67$ , than the control essays,  $M = 2.54, SD = .48$ . Finally, the authors of the affirmation essays referenced the self more frequently,  $F(1, 75) = 235.20, p < .001; M = 5.87, SD = 1.16$ , than control essays,  $M = 1.63, SD = 1.13$ . Taken together, these results indicate the self-affirmation manipulation was effective in not only priming participants to think about a cherished attribute of the self, but also affirmed participants' values and beliefs.

### Gender Effects

I subjected all variables to condition x gender ANOVAs and found a significant main effect of gender for the self-report distress composite score,  $F(1, 63) = 20.64, p < .01, partial \eta^2 = .25$ , but the interaction between gender and condition was not significant,  $F(2, 63) = 1.91, p =$

.16, *partial*  $\eta^2 = .06$ . The same pattern was observed for many of the individual variables (e.g., the total score of the TRGI). However, there was no main effect of gender on the behavioral distress composite score,  $F(1, 65) = .59, p = .45, \textit{partial} \eta^2 = .01$ , nor was the interaction significant,  $F(2, 65) = .23, p = .79, \textit{partial} \eta^2 = .01$ . These findings parallel those of Study 2, although in the present study, there was a main effect of gender on the self-report distress variables, rather than the interaction, yet the effect of gender was once again absent on the behavioral distress measure. As with Study 2, the self-report measures, but not the behavioral measure revealed this pattern. I therefore dropped gender from all remaining analyses. I will further address this finding in the Discussion.

### **Effect of Condition on Supplemental Measures (Self-Esteem and Affect)**

I calculated mean scores for the RSES ( $\alpha = .89$ ), and for both the positive ( $\alpha = .79$ ) and negative ( $\alpha = .89$ ) subscales of the PANAS. As with the previous two studies, these measures were merely included as filler items to mask the true purpose of the study, and are therefore not of primary interest. Once again, I did not have specific hypotheses concerning these items, however, a preliminary pattern seemed to be emerging such that there are generally no effects on self-esteem or positive affect, but the pattern of negative affect tends to mirror the other distress variables. In the interest of examining all aspects of the data as thoroughly as possible, I again conducted one-way between-subjects ANOVAs on the means of each in turn to examine whether the self-affirmation and killing manipulations influenced participants' feelings of self-worth or affect. Similar to the previous studies, there was no effect of condition on self-esteem,  $F(2, 71) = .20, p = .82, \textit{partial} \eta^2 = .01$ . Participants in the kill/no affirmation condition ( $M = 3.09, SD = .61$ ) reported the lowest levels of self-esteem compared to those in the kill/affirmation condition

( $M = 3.12$ ,  $SD = .43$ ), and the no kill/no affirmation condition ( $M = 3.18$ ,  $SD = .46$ ) was highest on this dimension, although these differences were non-significant.

In line with the previous results, there was no effect of condition on positive affect,  $F(2, 70) = 2.37$ ,  $p = .10$ , *partial*  $\eta^2 = .06$ , however, in contrast to previous findings, there was no effect on negative affect either,  $F(2, 72) = .25$ ,  $p = .78$ , *partial*  $\eta^2 = .01$ . Participants in the kill/affirmation condition reported slightly lower positive affect ( $M = 1.95$ ,  $SD = .48$ ) than those in the kill/no affirmation condition ( $M = 1.96$ ,  $SD = .54$ ), and participants in the no kill/no affirmation condition ( $M = 2.24$ ,  $SD = .55$ ) reported the highest positive affect overall, although this difference was not significant. Participants in the kill/affirmation condition also reported slightly higher negative affect ( $M = 1.92$ ,  $SD = .65$ ) than those in the kill/no affirmation condition ( $M = 1.81$ ,  $SD = .74$ ), and no kill/no affirmation condition ( $M = 1.78$ ,  $SD = .82$ ) reported the lowest negative affect overall, although again, these differences were non-significant.

### **Effect of Condition on Self-Report Distress**

I computed a distress composite score ( $\alpha = .89$ ) in the same manner as the previous studies to assess participants' overall guilt and distress on the self-report measures. I then subjected the distress score to a one-way between-subjects ANOVA to examine whether engaging in self-affirmation (or not) after killing (or not) affects distress. The results revealed that the effect of condition was not significant,  $F(2, 67) = .73$ ,  $p = .49$ , *partial*  $\eta^2 = .02$ . Participants in the kill/affirmation condition experienced more distress ( $M = .77$ ,  $SD = 5.17$ ) than those in the kill/no affirmation condition ( $M = -.89$ ,  $SD = 5.50$ ), and the no kill/no affirmation condition ( $M = -1.12$ ,  $SD = 7.02$ ) had the lowest levels of distress overall.

Although the distress composite was not significant, in the interest of further exploring the data and maintaining consistency across studies, I analyzed each distress variable individually to examine which aspects of distress may have been most influenced by the experimental manipulations (see Table 4 for means and test statistics). I subjected all variables to a one-way between-subjects ANOVA, and found no significant differences among any of the variables, including the mean of the total scale of the TRGI ( $\alpha = .89$ ) and the SSGS ( $\alpha = .92$ ). However, a marginally significant effect emerged for one of the subscales of the TRGI, namely Justification. Although non-significant, the pattern of results was generally the same across all variables, including the overall distress composite, such that participants in the kill/affirmation condition reported the highest distress compared to other two conditions, and those in the kill/no affirmation condition reported higher distress than the no kill condition.

Table 4

*Means, Standard Deviations, Cell Size, and Test Statistics for Each Individual Distress Variable from ANOVA on Self-Report Distress Composite, Including Total Scores for TRGI and SSGS (Study 3)*

Measure	Condition Means			Test Statistics
	Kill/Affirmation	Kill/No Affirmation	No Kill/No Affirmation	ANOVA
TRGI Total	2.58 (.55) <i>n</i> = 25	2.34 (.60) <i>n</i> = 23	2.29 (.77) <i>n</i> = 22	$F(2, 67) = 1.43$ , $p = .25$ , <i>partial</i> $\eta^2 = .04$
Distress	2.56 (.98) <i>n</i> = 26	2.23 (1.29) <i>n</i> = 24	2.13 (1.20) <i>n</i> = 24	$F(2, 71) = .96$ , $p = .39$ , <i>partial</i> $\eta^2 = .03$
Guilt Cognitions	2.63 (.51) <i>n</i> = 25	2.44 (.54) <i>n</i> = 24	2.34 (.72) <i>n</i> = 23	$F(2, 69) = 1.40$ , $p = .25$ , <i>partial</i> $\eta^2 = .04$
Hindsight	2.61 (.73) <i>n</i> = 26	2.65 (.67) <i>n</i> = 24	2.58 (.84) <i>n</i> = 24	$F(2, 71) = .06$ , $p = .94$ , <i>partial</i> $\eta^2 = .00$
Wrongdoing	1.99 (.60) <i>n</i> = 25	1.66 (.52) <i>n</i> = 25	1.74 (.62) <i>n</i> = 22	$F(2, 69) = 2.21$ , $p = .12$ , <i>partial</i> $\eta^2 = .06$

Justification	3.41 (.77) <i>n</i> = 26	3.19 (.99) <i>n</i> = 25	2.80 (1.14) <i>n</i> = 24	$F(2, 72) = 2.49$ , $p = .09$ , <i>partial</i> $\eta^2 = .07$
SSGS Total	2.51 (.71) <i>n</i> = 26	2.49 (.79) <i>n</i> = 25	2.44 (.78) <i>n</i> = 24	$F(2, 72) = .06$ , $p$ $= .94$ , <i>partial</i> $\eta^2 = .00$
Shame	1.58 (.76) <i>n</i> = 26	1.55 (.71) <i>n</i> = 25	1.70 (.75) <i>n</i> = 24	$F(2, 72) = .27$ , $p$ $= .77$ , <i>partial</i> $\eta^2 = .01$
Guilt	2.15 (1.08) <i>n</i> = 26	2.18 (1.23) <i>n</i> = 25	2.00 (1.16) <i>n</i> = 24	$F(2, 72) = .18$ , $p$ $= .84$ , <i>partial</i> $\eta^2 = .01$
Pride	3.78 (.71) <i>n</i> = 26	3.74 (.88) <i>n</i> = 25	3.61 (.72) <i>n</i> = 24	$F(2, 72) = .34$ , $p$ $= .71$ , <i>partial</i> $\eta^2 = .01$

*Note.* Standard deviations appear in parentheses

### Effect of Condition on Behavioral Guilt (Hand Washing)

I computed a distress composite score for hand washing in the same way as Study 2 to assess the overall behavioral guilt and distress. I then subjected this hand washing composite score ( $\alpha = .45$ ) to a one-way between-subjects ANOVA, which revealed a significant main effect of condition,  $F(2, 69) = 6.15$ ,  $p < .01$ , *partial*  $\eta^2 = .15$ . The pattern suggests participants in the kill/no affirmation condition ( $M = .22$ ,  $SD = 1.91$ ) displayed higher levels of guilt and distress, compared to participants in the kill/affirmation condition ( $M = .07$ ,  $SD = 2.06$ ). The no kill/no affirmation condition ( $M = -1.42$ ,  $SD = 1.17$ ) displayed the lowest levels overall.

Because there were three conditions and because I had a-priori predictions, I performed a series of planned contrasts. The hypothesized differences among the conditions were that kill/affirmation would lead to the highest amount of guilt/distress, followed by kill/no affirmation, and then no kill/no affirmation. To assess this predicted pattern, I conducted three planned contrasts. Contrast 1 compared just the two kill conditions (kill/affirmation vs. kill/no affirmation) to see if they differed from each other (as they had in the previous two studies), while ignoring the no kill (no affirmation) condition. Contrast 2 compared the kill/affirmation to

the no kill (no affirmation) condition to determine whether killing and engaging in self-affirmation differed from not killing, while ignoring kill/no affirmation. Contrast 3 examined the two no affirmation conditions (kill/no affirmation and no kill/no affirmation), while ignoring the kill/affirmation condition. Contrasts 2 and 3 were significant, indicating both kill conditions differed from not killing, but did not differ from each other (i.e., self-affirmation did not differ from no affirmation after killing). See Table 6 for the results of all planned comparisons.

I then analyzed each of the hand washing variables separately to examine which aspects of behavioral distress may have been most influenced by the self-affirmation and killing manipulations. See Table 5 for means and statistics for all variables, including both the raw (mean amount of time, mean amount of soap, and mean number of paper towels) as well as the transformed  $z$  scores. I subjected all variables to a one-way between-subjects ANOVA, and a significant effect was found for the number of paper towels used, however, effects for both the amount of time spent (in seconds) and the amount of soap used (in grams) were both non-significant on their own. In general, the pattern of results suggests participants in the kill/no affirmation condition displayed the highest level of distress followed by the kill/affirmation condition, with the no kill/no affirmation condition exhibiting the least distress overall (see Table 5).

As with the overall behavioral distress composite, I performed the same three planned contrasts on each of the individual distress measures. Results of the planned contrasts appear in Table 6. In general, the same pattern found on the overall composite was also observed on the individual measures: killing the bugs led to more distress than not killing, but self-affirmation vs. no affirmation did not affect overall distress.

Table 5

*Means, Standard Deviations, Cell Size, and Test Statistics for Each Individual Distress Variable from ANOVA on Behavioral Distress Composite, Including Both Raw and Standardized Z Scores for Time, Towels, and Soap (Study 3)*

Measure	Condition Means			Test Statistics
	Kill/Affirmation	Kill/No Affirmation	No Kill/No Affirmation	ANOVA
Time	17.04 (8.37) <i>n</i> = 26	17.74 (6.61) <i>n</i> = 23	13.78 (5.98) <i>n</i> = 23	$F(2, 69) = 2.04, p = .14,$ $partial \eta^2 = .06$
Towels	1.69 (.74) <i>n</i> = 26	1.74 (.62) <i>n</i> = 23	1.25 (.44) <i>n</i> = 24	$F(2, 70) = 4.63, p = .01,$ $partial \eta^2 = .12$
Soap	.77 (.44) <i>n</i> = 26	.77 (.49) <i>n</i> = 23	.59 (.15) <i>n</i> = 24	$F(2, 70) = 1.86, p = .16,$ $partial \eta^2 = .05$
Time <i>z</i> Scores	-.08 (.85)	-.01 (.67)	-.41 (.61)	
Towels <i>z</i> Scores	.11 (1.19)	.19 (1.00)	-.61 (1.04)	
Soap <i>z</i> Scores	.04 (1.03)	.05 (1.16)	-.40 (.35)	

*Note.* Standard deviations appear in parentheses.

Table 6

*Planned Contrasts of Kill/Affirmation, Kill/No Affirmation, and No Kill/No Affirmation Conditions on All Behavioral Distress Measures (Study 3)*

Measure	Test Statistics		
	Kill/Affirmation vs. Kill/No Affirmation	Kill/Affirmation vs. No Kill/No Affirmation	Kill/No Affirmation vs. No Kill/No Affirmation
Time	$t(69) = -.34, p = .73$	$t(69) = 1.60, p = .12$	$t(69) = 1.88, p = .06$
Towels	$t(70) = -.24, p = .81$	$t(70) = 2.60, p = .01$	$t(70) = 3.11, p < .001$
Soap	$t(70) = -.02, p = .98$	$t(70) = 2.04, p = .05$	$t(70) = 1.76, p = .09$
Composite	$t(69) = -.27, p = .79$	$t(69) = 3.16, p < .001$	$t(69) = 3.53, p = .001$

## Discussion

The results of Study 3 offered partial support for the idea that self-affirmation would increase the distress of killing by highlighting the discrepancy between self-standards and behavior. Whereas Study 2 showed self-affirmation led to an increase in guilt and distress, Study 3 showed self-affirmation had little or no effect on the distress of killing. In other words, killing the bugs increased participants' distress above a no kill control condition, but self-affirmation had no effect on increasing or decreasing distress once they had already killed. On the one hand, this finding supports the overall analysis that self-affirmation, a procedure that typically reduces dissonance and distress, failed to buffer the distress of killing. On the other hand, the results of Study 3 did not support the strong version of my hypothesis, namely that self-affirmation would exacerbate killing-related distress.

These results were somewhat surprising. However, as noted, participants in this study completed the writing (self-affirmation) manipulation after the extermination task, whereas in the previous studies they completed the self-affirmation task prior to the extermination. The difference in findings between Study 3 and Study 2 might therefore be attributable to this alteration in design, since this is the primary change made between the studies. It appears the timing of the self-affirmation intervention may be an important factor in whether or not it will affect distress. Although this explanation is plausible, it is also worth noting that I had to make a large number of exclusions in the present study, particularly in the no kill control condition. Given that the pattern of results between Studies 2 and 3 is not drastically different (killing still led to more distress than not killing), it is possible that the number of exclusions are responsible for the present findings, rather than the timing of the manipulation. Moreover, it is important to

note that although self-affirmation did not increase guilt and distress over no affirmation, self-affirmation also did not reduce distress, as the bulk of research on self-affirmation would predict.

Another surprising (and somewhat more concerning) finding was that no effects were obtained on the self-report measures of distress. This revelation is in stark contrast to what has been observed in the prior two studies, both of which found strong effects for the manipulation on self-reported levels of distress (albeit in different directions). This lack of movement on the self-report items is puzzling. However, effects did emerge on the behavioral measure using the same hand-washing task as in Study 2. Taken together, the results of Studies 2 and 3 suggest the behavioral measure is more sensitive to actual differences in guilt and distress than the self-report measures. For example, in both Study 2 and 3, gender effects occurred on the self-report measures, but not on the behavioral measure. This might indicate participants are cognizant of what the self-report measures are assessing and are attempting to mask their true feelings, whereas because the hand-washing task is more inconspicuous and they are unaware of what this task assesses, they are less able to inhibit their distress on this implicit measure.

Along the same lines, there was once again a gender effect on the self-report measures, although this time it was a main effect, rather than the interaction discovered in Study 2. However, as with Study 2, it appears that this finding seems essentially to be an effect of men not wanting to admit on the self-report measures that they were distressed by the extermination task. Yet the behavioral measure seems to more accurately capture the distress, as participants do not connect the hand-washing task with the rest of the study, and men appear as equally distressed as their female counterparts did. Additionally, the cell sizes for the distribution of men vs. women across the conditions are somewhat better than those in Study 2, although still less than ideal (e.g., there are less than 10 men in some conditions). Taken together, given that gender effects

do not show up on the behavioral measures, the results for gender should be interpreted with caution.

Although the results suggest affirming the self after the act of killing has no effect on guilt/distress, there is at least one limitation in the present study that should be noted. The no kill control condition in particular presented some challenges. Despite giving participants explicit instructions that the grinder tube does not lead into the machine, making it impossible to harm or kill the bugs, they still expressed remorse or distress after completing the task. In terms of the exclusions that had to be made, two participants in the no affirmation condition chose not to complete the study (compared to zero in the affirmation condition), whereas nine participants in the no kill condition were excluded due to misunderstanding they were not killing the bugs. Many participants in this condition revealed during the debriefing that they still felt bad about what they had done and were bothered by the procedure. Other participants seemed to understand they had not actually killed the bugs, but still seemed to be concerned they had somehow hurt the bugs just by placing them in the grinder (and pushing the activation button, which, admittedly, makes the machine vibrate and is rather noisy). Due to this unexpected complication (of the neutral, no kill control condition not serving as a neutral control as intended), quite a few participants had to be excluded from the present study. Thus, although participants in the no kill condition did experience less distress overall compared to the kill conditions, participants in this condition still seemed to experience some distress.

In light of the limitations of the present study, a fourth study was necessary to assess whether the self-affirmation task increases distress when it comes before the act of killing, as observed in Study 2. Although Study 2 supported the strong version of my hypothesis, it lacked a no kill control condition. Given the lack of a no kill condition in Study 2, and the issues with

this condition in Study 3, it remains possible that the killing (no affirmation) condition did not increase distress over and above the no kill control condition. I thus included a no kill condition (again) in Study 4. If the different pattern of results between Studies 2 and 3 is indeed due to the change in the timing of the self-affirmation intervention, then moving the manipulation before the extermination task again (as in Studies 1 and 2) should account for this difference and allow for a more direct replication of Study 2. However, given the issues encountered in Study 3, I modified the no kill condition so participants would better understand the procedure would not harm the bugs in any way.

#### **Study 4**

Study 4 was designed to provide a more direct replication of Study 2 with the inclusion of the additional (no kill) control condition introduced in Study 3, with a slight modification. Participants were therefore randomly assigned to one of three conditions and all participants were asked to complete the extermination task and kill a standardized number of (10) bugs. The first two conditions (kill/affirmation vs. kill/no affirmation) were identical to those used in the previous three studies, in which participants were all led to believe they were killing bugs. In the third (no kill/no affirmation) condition, I directly informed participants that they would not be killing the bugs; instead they would pretend to kill the bugs by miming the procedure (i.e., they did not dump the bugs into the grinder at all). This condition was designed to be a more neutral control condition compared to Study 3, in that participants knew they were not actually harming the insects, and therefore allowed for cleaner comparisons between the conditions (rather than just between self-affirmation and no affirmation, but also between killing and not killing). The idea behind this modification was to account for some of the issues that arose in Study 3 in which participants seemed to have qualms about placing the bugs in the grinder in the no kill

condition out of fear they were still somehow causing the bugs harm or distress. By removing that step and never having the bugs enter the grinder, any ambiguity the participants might have had about harming the bugs should have been eliminated.

In line with Studies 2 and 3, I predicted a linear pattern of results, such that the kill/affirmation condition would lead to the highest levels of guilt and distress after killing the bugs, followed by the kill/no affirmation condition, and the no kill/no affirmation condition would display the least distress. To assess this hypothesis, all participants first completed the self-affirmation task, followed by the extermination task, and the measures of guilt and distress (including the behavioral hand washing measure). This design enabled me to determine if self-affirmation under these circumstances (of killing 10 bugs) is indeed more distressing, and if the timing of the self-affirmation manipulation is important, as suggested by the results of Study 3.

## **Method**

### **Participants**

One hundred nineteen undergraduates (37 men, 79 women, 3 unreported) from the University of Alberta participated in exchange for partial credit towards the completion of an introductory psychology course. Participants were at least 18 years of age ( $M = 18.90$ ,  $SD = 1.54$ ). However, several participants were excluded from analyses due to suspicion regarding the tasks (e.g., did not believe they were actually killing the bugs) and/or research hypotheses, not following instructions, or experimenter error. A final sample of 91 (32 men, 57 women, 2 unreported) participants were included in all analyses for which they had complete data ( $M_{age} = 18.95$ ,  $SD_{age} = 1.63$ ). All requirements established by the Ethics Review Board at the University of Alberta were satisfied, as well as the APA ethical standards for participation with human research subjects.

## Research Design

Study 4 utilized a one-way between-subjects design, manipulating one independent variable. As with Study 3, there were three conditions. Participants were randomly assigned to either the kill/affirmation condition ( $n = 30$ ), kill/no affirmation ( $n = 33$ ), or no kill/no affirmation control ( $n = 28$ ), and the dependent variables included several self-report measures of distress (e.g., guilt and shame) as well as a behavioral measure of distress. Once again, Study 4 employed a double-blind procedure.

## Materials and Procedure

The materials and procedure for Study 4 were largely the same as those of Study 3. All participants completed the study individually. Upon arrival, the experimenter greeted participants in the hallway and brought them into the laboratory where they completed an informed consent form after a brief introduction to the study. I used the same cover story from the previous studies, such that participants were told the study was designed to investigate personality and exterminators. The researcher told participants the study involved completing a bug extermination task as well as answering some questions regarding their experience. Participants then completed a series of tasks independently, at their own pace.

As with Studies 1 and 2 (but not Study 3), all participants first completed a packet of “personality measures” containing the experimental manipulation (i.e., either self-affirmation or no affirmation control) to which they had been randomly assigned. Next, participants were ushered by the researcher from the main room to the “extermination area” in an adjacent room to complete the extermination task. I again employed the same set-up as the previous studies. Specifically, the “extermination area” contained a desk and all the materials required to complete the task. The experimenter instructed the participants to: (1) read through the information

packet, (2) watch the video on the laptop, (3) complete the extermination task as demonstrated in the video (in which they were either supposed to kill – or mime killing – the 10 pill bugs), and (4) return to the main room for further instructions.

As noted, I also modified the no kill control condition for the present study from the method used in Study 3, as an attempt to ensure participants fully understood they were not harming the insects in any way, in order to address some of the issues that arose in the previous study. As with Study 3, the written instructions at the end of the demonstration video stated no bugs would be harmed through the procedures of the study. In addition, the video instructed participants to “mime” the procedure. Rather than dumping the insects into the tube on the grinder, the video instead instructed participants in this condition to lift each cup containing each bug to the grinder, but importantly, to then return the cup to the table without dumping the bug into the tube. After they had mimed the procedure with each of the bugs, they were then to push the activation button on the top of the grinder to finish simulating the extermination task. All other components of the study (for participants in the two kill conditions – self-affirmation and no affirmation) remained the same as in Study 2. Thus, all participants were instructed to kill (or, for the no kill condition, mime killing) 10 bugs.

As with the prior two studies, participants were also provided with an opportunity to wash their hands after they returned from the extermination area, as a measure of behavioral guilt. Once participants had finished washing their hands, the researcher opened the door and entered the main room with the final packet of questionnaires containing, in order, the TRGI, PANAS, and SSGS. After explaining the final packet, the researcher then left the room so the participant could complete the final packet in private. Upon completion of all tasks, the researcher probed participants for suspicion and thoroughly debriefed them.

## Results

### Manipulation Checks

To assess whether the self-affirmation manipulation worked as intended, three independent raters read each essay and coded the responses on the same four questions used in the previous studies. I then computed Intraclass correlations (ICCs) across the three raters for each of the four questions. The ICCs all indicated excellent levels of agreement (Fleiss, 1981) between raters for all four items (two-way mixed effects model, absolute agreement, all ICCS > .92). I then conducted one-way between-subjects ANOVAs on each of the four items to determine whether self-affirmation led to higher ratings, as predicted. The ANOVAs revealed participants in the affirmation condition indicated the value was more important to the participant,  $F(1, 91) = 733.61, p < .001; M = 6.47, SD = .61$ , compared to those in the control condition,  $M = 2.21, SD = .75$ . The affirmation essays were also rated as being more affirming to the participants' beliefs,  $F(1, 91) = 507.24, p < .001; M = 6.04, SD = .79$ , versus  $M = 2.53, SD = .65$ , as well as increasing participants' confidence in their value as a person,  $F(1, 91) = 472.27, p < .001; M = 5.63, SD = .98$ , than the control essays,  $M = 2.27, SD = .50$ . Finally, the authors of the affirmation essays referenced the self more frequently,  $F(1, 91) = 216.43, p < .001; M = 5.68, SD = 1.24$ , than control essays,  $M = 1.73, SD = 1.18$ . Taken together, these results indicate the self-affirmation manipulation was effective in not only priming participants to think about a cherished attribute of the self, but also affirmed participants' values and beliefs.

### Gender Effects

I subjected all variables to condition x gender ANOVAs and found a significant main effect of gender for the self-report distress composite score,  $F(1, 77) = 5.79, p = .02, partial \eta^2 = .07$ , and the interaction between gender and condition was non-significant, but marginal,  $F(2, 77)$

= 2.74,  $p = .07$ , *partial*  $\eta^2 = .07$ . The same pattern was observed for some of the individual variables (e.g., the total score of the SSGS). However, there was no main effect of gender on the behavioral distress composite score,  $F(1, 82) = .82$ ,  $p = .37$ , *partial*  $\eta^2 = .01$ , nor was the interaction significant,  $F(2, 82) = .48$ ,  $p = .62$ , *partial*  $\eta^2 = .01$ . These findings parallel those of Study 3 (and the effect of gender is even weaker here), such that the main effect of gender is only present on some of the self-report measures (and on fewer than in Study 3), but is absent on the behavioral distress measure. I therefore dropped gender from all remaining analyses. I will further address this finding in the Discussion.

### **Effect of Condition on Supplemental Measures (Self-Esteem and Affect)**

I computed mean scores for the RSES ( $\alpha = .91$ ), and for both the positive ( $\alpha = .84$ ) and negative ( $\alpha = .89$ ) subscales of the PANAS. As with the previous studies, these measures were simply included as filler items designed to bolster the cover study, and are therefore not of primary interest. Once again, I did not have specific hypotheses concerning these items, however, the general pattern seems to be that there are usually no effects on self-esteem or positive affect, but negative affect tends to mirror the pattern observed on the other distress variables. In the interest of thoroughly examining all facets of the data (as well as maintaining consistency with the previous studies), I again conducted one-way between-subjects ANOVAs on the means of each in turn to examine whether the self-affirmation and killing manipulations affected participants' feelings of self-worth or affect. There was no effect of condition on self-esteem,  $F(2, 87) = .82$ ,  $p = .46$ , *partial*  $\eta^2 = .02$ . Participants in the no kill/no affirmation condition reported the highest levels of self-esteem ( $M = 3.31$ ,  $SD = .46$ ), followed by the kill/no affirmation condition ( $M = 3.27$ ,  $SD = .52$ ), and the kill/affirmation condition reported slightly lower self-esteem ( $M = 3.13$ ,  $SD = .70$ ), although these differences were non-significant.

In slight contrast to the previous studies, there was a main effect of condition on positive affect,  $F(2, 88) = 3.89, p = .02, \text{partial } \eta^2 = .08$ , however, the main effect of condition on negative affect was non-significant,  $F(2, 88) = .93, p = .40, \text{partial } \eta^2 = .02$ . Participants in the kill/affirmation condition reported lower positive affect ( $M = 1.85, SD = .63$ ) than those in the kill/no affirmation condition ( $M = 1.95, SD = .58$ ), and participants in the no kill/no affirmation condition had the highest levels of positive affect ( $M = 2.29, SD = .69$ ). Participants in the kill/affirmation condition also reported higher negative affect ( $M = 2.00, SD = .83$ ) than those in the kill/no affirmation condition ( $M = 1.83, SD = .88$ ), and the no kill/no affirmation condition had the lowest negative affect ( $M = 1.72, SD = .65$ ), though once again, these differences were not significant.

### **Effect of Condition on Self-Report Distress**

I calculated the same distress composite score ( $\alpha = .90$ ) used in the previous studies to assess participants' overall guilt and distress on the self-report measures. I then subjected the distress score to a one-way between-subjects ANOVA to examine whether self-affirmation (or not) and killing (or not) affects distress. The results revealed a significant main effect of condition,  $F(2, 82) = 9.53, p < .001, \text{partial } \eta^2 = .19$ . In line with the hypothesis (and the results of Study 2), participants in the kill/affirmation condition experienced more distress ( $M = 2.36, SD = 5.91$ ) than those in the kill/no affirmation condition ( $M = 1.79, SD = 5.69$ ), and the no kill/no affirmation ( $M = -3.31, SD = 4.08$ ) experienced the least distress overall.

I then analyzed each distress variable separately to examine which aspects of distress may have been most influenced by the self-affirmation and extermination task manipulations. See Table 7 for means and statistics for all variables. One-way between-subjects ANOVAs revealed significant differences on nine of the variables, including the total scale of the TRGI ( $\alpha$

= .91), and the Distress, Guilt Cognitions, Hindsight, and Justification subscales, as well as the total scale of the SSGS ( $\alpha = .93$ ), and the three subscales (Shame, Guilt, and Pride).

Additionally, I found a marginally significant effect for the mean of the remaining subscale of the TRGI, namely Wrongdoing. Notably, the pattern of results was generally the same for each variable (regardless of significance level) as the overall distress composite, such that the kill/affirmation condition reported the highest levels of distress, followed by the kill/no affirmation, and the no kill/no affirmation was lowest. However, there were a couple of instances where the kill/affirmation and kill/no affirmation participants scored equally high on a given distress measure (e.g., Hindsight and Wrongdoing). Given that there was a significant main effect, follow-up tests were needed to determine which conditions differed significantly from each other.

Following Study 3, because there was a significant main effect of condition and specific predictions concerning the three conditions, I performed the same series of (three) planned comparisons on each variable (including the overall distress composite) to further probe the pattern of results. In general, the pattern suggests that Contrasts 2 and 3 are significant across the majority of the measures (see Table 8 for all comparisons on all self-report variables), indicating both kill conditions (affirmation vs. no affirmation) are different from the no kill (no affirmation) condition, however, the two kill conditions do not differ from each other.

Table 7

*Means, Standard Deviations, Cell Size, and Test Statistics for Each Individual Distress Variable from ANOVA on Self-Report Distress Composite, Including Total Scores for TRGI and SSGS (Study 4)*

Measure	Condition Means			Test Statistics
	Kill/Affirmation	Kill/No Affirmation	No Kill/No Affirmation	ANOVA
TRGI Total	2.72 (.73) <i>n</i> = 30	2.68 (.68) <i>n</i> = 30	2.07 (.55) <i>n</i> = 27	$F(2, 84) = 8.44$ , $p < .001$ , <i>partial</i> $\eta^2 = .17$
Distress	2.67 (1.45) <i>n</i> = 30	2.48 (1.14) <i>n</i> = 32	1.67 (.68) <i>n</i> = 28	$F(2, 87) = 6.16$ , $p < .01$ , <i>partial</i> $\eta^2 = .12$
Guilt Cognitions	2.77 (.65) <i>n</i> = 30	2.73 (.64) <i>n</i> = 31	2.18 (.56) <i>n</i> = 27	$F(2, 85) = 7.83$ , $p = .001$ , <i>partial</i> $\eta^2 = .16$
Hindsight	2.86 (.80) <i>n</i> = 30	2.87 (.83) <i>n</i> = 33	2.32 (.77) <i>n</i> = 27	$F(2, 87) = 4.43$ , $p = .02$ , <i>partial</i> $\eta^2 = .09$
Wrongdoing	1.96 (.75) <i>n</i> = 30	1.95 (.71) <i>n</i> = 32	1.59 (.65) <i>n</i> = 28	$F(2, 87) = 2.62$ , $p = .08$ , <i>partial</i> $\eta^2 = .06$
Justification	3.68 (.96) <i>n</i> = 30	3.47 (.90) <i>n</i> = 32	2.90 (.97) <i>n</i> = 28	$F(2, 87) = 5.20$ , $p = .01$ , <i>partial</i> $\eta^2 = .11$
SSGS Total	2.75 (.85) <i>n</i> = 28	2.60 (.80) <i>n</i> = 33	1.99 (.52) <i>n</i> = 28	$F(2, 86) = 8.40$ , $p < .001$ , <i>partial</i> $\eta^2 = .16$
Shame	1.79 (.87) <i>n</i> = 29	1.69 (.94) <i>n</i> = 33	1.26 (.38) <i>n</i> = 28	$F(2, 87) = 3.62$ , $p = .03$ , <i>partial</i> $\eta^2 = .08$
Guilt	2.49 (1.35) <i>n</i> = 30	2.28 (1.10) <i>n</i> = 33	1.54 (.70) <i>n</i> = 28	$F(2, 88) = 6.00$ , $p < .01$ , <i>partial</i> $\eta^2 = .12$
Pride	4.00 (.76) <i>n</i> = 29	3.84 (.80) <i>n</i> = 33	3.16 (.96) <i>n</i> = 28	$F(2, 87) = 7.97$ , $p = .001$ , <i>partial</i> $\eta^2 = .16$

*Note.* Standard deviations appear in parentheses.

Table 8

*Planned Contrasts of Kill/Affirmation, Kill/No Affirmation, and No Kill/No Affirmation Conditions on All Self-Report Distress Measures (Study 4)*

Measure	Test Statistics		
	Kill/Affirmation vs. Kill/No Affirmation	Kill/Affirmation vs. No Kill/No Affirmation	Kill/No Affirmation vs. No Kill/No Affirmation
TRGI Total	$t(84) = .26, p = .80$	$t(84) = 3.71, p < .001$	$t(84) = 3.46, p = .001$
Distress	$t(87) = .57, p = .57$	$t(87) = 3.38, p < .01$	$t(87) = 3.37, p = .001$
Guilt Cognitions	$t(85) = .24, p = .81$	$t(85) = 3.56, p = .001$	$t(85) = 3.35, p = .001$
Hindsight	$t(87) = -.04, p = .97$	$t(87) = 2.56, p = .01$	$t(87) = 2.66, p = .01$
Wrongdoing	$t(87) = .06, p = .96$	$t(87) = 2.01, p = .05$	$t(87) = 1.99, p = .05$
Justification	$t(87) = .86, p = .39$	$t(87) = 3.12, p < .01$	$t(87) = 2.33, p = .02$
SSGS Total	$t(86) = .70, p = .49$	$t(86) = 4.04, p < .001$	$t(86) = 3.60, p = .001$
Shame	$t(87) = .42, p = .68$	$t(87) = 2.97, p = .01$	$t(87) = 2.39, p = .02$
Guilt	$t(88) = .67, p = .51$	$t(88) = 3.38, p < .01$	$t(88) = 3.17, p < .01$
Pride	$t(87) = .74, p = .46$	$t(87) = 3.75, p < .001$	$t(87) = 3.14, p < .01$
Distress Composite	$t(82) = .41, p = .68$	$t(82) = 3.96, p < .001$	$t(82) = 3.62, p = .001$

### **Effect of Condition on Behavioral Guilt (Hand Washing)**

I created the same distress composite score for hand washing as in the previous two studies to assess overall behavioral guilt and distress. I then subjected this hand washing composite score ( $\alpha = .35$ ) to a one-way between-subjects ANOVA, which revealed a significant main effect of condition,  $F(2, 87) = 3.54, p = .03, partial \eta^2 = .08$ . Consistent with what was observed on the self-report measures, participants in the kill/affirmation condition ( $M = .14, SD = 1.72$ ) displayed higher levels of guilt and distress, compared to participants in the kill/no affirmation condition ( $M = -.61, SD = 1.43$ ), and the no kill/no affirmation condition ( $M = -.85, SD = 1.26$ ) was the lowest overall.

I then analyzed each of the hand washing variables separately to examine which aspects of behavioral distress may have been most influenced by the experimental manipulations. See Table 9 for means and statistics for all variables, including both the raw (mean amount of time,

mean amount of soap, and mean number of paper towels) as well as the transformed  $z$  scores. I subjected all variables to one-way between-subjects ANOVAs, revealing significant effects for the amount of soap used (in grams); however, effects for both the amount of time spent (in seconds) and the number of paper towels used were non-significant. In general, the pattern suggests participants in the kill/affirmation condition displayed the highest levels of distress, followed by those in the kill/no affirmation condition and the no kill/no affirmation condition had the lowest levels overall. Although when looking at the number of paper towels used, participants in the no kill condition used slightly more than those in the kill/no affirmation condition (see Table 9).

Because the main effect of condition was significant, I performed the same three planned contrasts on the behavioral distress composite as well as the individual hand washing variables, in line with what was done with the self-report distress composite. See Table 10 for all planned comparisons. For the overall behavioral distress composite, contrasts 1 and 2 were significant. The next best individual indicator was the amount of soap used, which followed a similar pattern, with contrasts 2 and 3 reaching significance, and contrast 1 was marginally significant. Taken together, the results suggest the kill/affirmation condition experienced significantly higher distress than both the kill/no affirmation and the no kill/no affirmation conditions. Additionally, killing (with both affirmation and no affirmation) was significantly more distressing than not killing. However, one interesting finding is the kill/no affirmation condition does not appear to differ significantly from the no kill/no affirmation condition, except when looking at the amount of soap used (see Table 10 for results of all planned contrasts).

Table 9

*Means, Standard Deviations, Cell Size, and Test Statistics for Each Individual Distress Variable from ANOVA on Behavioral Distress Composite, Including Both Raw and Standardized Z Scores for Time, Towels, and Soap (Study 4)*

Measure	Condition Means			Test Statistics
	Kill/Affirmation	Kill/No Affirmation	No Kill/No Affirmation	ANOVA
Time	16.93 (6.66) <i>n</i> = 30	15.69 (6.08) <i>n</i> = 32	15.50 (6.03) <i>n</i> = 28	$F(2, 87) = .46, p = .63, \text{partial } \eta^2 = .01$
Towels	1.87 (.78) <i>n</i> = 30	1.63 (.61) <i>n</i> = 32	1.68 (.61) <i>n</i> = 28	$F(2, 87) = 1.09, p = .34, \text{partial } \eta^2 = .03$
Soap	.75 (.33) <i>n</i> = 30	.63 (.22) <i>n</i> = 32	.50 (.22) <i>n</i> = 28	$F(2, 87) = 6.70, p = .002, \text{partial } \eta^2 = .13$
Time <i>z</i> Scores	-.10 (.77)	-.24 (.70)	-.26 (.69)	
Towels <i>z</i> Scores	.15 (1.11)	-.20 (.87)	-.12 (.87)	
Soap <i>z</i> Scores	.09 (.72)	-.17 (.49)	-.46 (.48)	

*Note.* Standard deviations appear in parentheses.

Table 10

*Planned Contrasts of Kill/Affirmation, Kill/No Affirmation, and No Kill/No Affirmation Conditions on All Behavioral Distress Measures (Study 4)*

Measure	Test Statistics		
	Kill/Affirmation vs. Kill/No Affirmation	Kill/Affirmation vs. No Kill/No Affirmation	Kill/No Affirmation vs. No Kill/No Affirmation
Time	$t(87) = .78, p = .44$	$t(87) = .87, p = .39$	$t(87) = .12, p = .91$
Towels	$t(87) = 1.42, p = .16$	$t(87) = 1.07, p = .29$	$t(87) = -.31, p = .76$
Soap	$t(87) = 1.80, p = .08$	$t(87) = 3.66, p < .001$	$t(87) = 1.95, p = .05$
Composite	$t(87) = 1.99, p = .05$	$t(87) = 2.53, p = .01$	$t(87) = .61, p = .54$

## Discussion

The results of Study 4 provide support for the hypothesis that self-affirmation can exacerbate distress after killing. In line with this hypothesis, the present study replicated the key findings of Study 2, such that self-affirmation led to more guilt and distress after killing. Although the difference between the kill/affirmation and kill/no affirmation condition was not significant on the self-report measures, the two conditions did differ significantly from each other on the behavioral measure. The self-report measures therefore offered support for the weak version of my hypothesis, namely self-affirmation should have little or no effect on reducing the distress of killing. However, the behavioral hand washing measure showed support for the strong version of my hypothesis, that self-affirmation would increase killing-related distress. Participants in the kill/affirmation condition exhibited more behavioral distress than did participants in the other two conditions. As noted, the behavioral hand washing measure may be a better indication of participants' true levels of distress given that it is subtle and thus less influenced by self-serving motives or self-presentation motives to appear less distressed.

Once again, the various manipulations did not affect self-esteem. However, one minor difference between the current study and the previous ones is there was an effect on positive affect, but not negative affect. Although this effect was somewhat surprising, it fits with the observed pattern on the other dependent measures, such that participants in the kill/affirmation condition experienced lower positive affect than those in the other two conditions. Moreover, even though the effect on negative mood was not significant, the pattern fit what would be expected, such that participants in the kill/affirmation condition experienced higher levels of negative affect.

One unexpected result was the kill/no affirmation condition did not appear to differ from the no kill condition on the behavioral distress measure. This finding is surprising and it is unclear exactly why this would be the case as killing should be more distressing than not. Unlike with some of the limitations of the previous studies, this finding is perhaps not directly attributable to changes in the design. As with Study 2, participants completed the affirmation manipulations prior to the extermination task, and the instructions for this task remained constant across all four studies. However, it is important to note that the pattern of results shows the highest level of distress was experienced by those who engaged in self-affirmation prior to killing, followed by those who also killed, but did not engage in self-affirmation, and the lowest levels of distress, as hypothesized, were experienced by those in the control (no kill) condition. This distribution of means across conditions is also in line with Study 2, giving credence to the notion that the timing of the affirmation intervention is potentially important.

Although there may not have been any issues with the design that are readily apparent, one concern worth noting is many participants were excluded from the present study, much more so than any of the previous studies, and there was also a much higher rate of attrition. For some reason, a large number of participants refused to complete the extermination task, either immediately upon receiving information they would be killing bugs, or at some point in the procedure (e.g., usually after watching the demonstration video). It is not clear why this occurred, as the demonstration video, materials, and procedure were largely the same as in Studies 1 through 3, with the exception of the discussed changes to instructions on the demonstration video. For example, in terms of attrition, 11 participants in the self-affirmation condition, as well as 11 in the no affirmation condition bowed out of the study before completing the extermination task, compared to zero in the no kill condition. Moreover, when compared to

prior studies (e.g., Study 3 in which zero participants in the affirmation condition opted out when the affirmation task came after killing), these numbers are startling.

One explanation for the pattern of findings may therefore relate to the levels of attrition in the present study. It is possible the participants who opted-out of the study were those who would have been most distressed by the procedure (and indeed that is the reason they did not complete the bug-killing task). This self-selected attrition may have left a sample of participants who were less prone to be distressed by the procedure, leading to lower levels of distress in the kill/no affirmation condition than the levels of distressed observed previously. Moreover, although a similar number of participants opted-out of the kill/affirmation condition (and thus those remaining may have been more comfortable with the bug-killing task), the self-affirmation task still had the power to make participants feel bad about what they had done and increased their distress as a result.

Another finding worthy of mention is the main effect of gender was once again observed on some of the self-report measures of distress. Similar to Study 3, however, the interaction between gender and condition failed to reach significance. Importantly, the gender effect appears markedly weaker in the present study than it was in the previous two (e.g., fewer measures showed the effect compared to Study 3). One potential explanation for this finding is the sample size of the present study was larger than in the previous studies, averaging roughly 30 participants per condition. As noted in previous discussions of this effect, many of the cells with men across conditions are significantly small, calling the validity of this finding into question. Given that the present study had a larger sample (including more men than those of the previous studies), yet the finding was significantly weaker, suggests the gender effects may be attributable to an artifact of sample size. I believe if we were to collect data on more participants (and

specifically, to even out the number of men and women across conditions), the gender effect would likely disappear.

Even more compelling, the main effect of gender was once again only apparent on the self-report measures, but was lacking on the behavioral distress measure. This same pattern was observed across all three studies employing the hand washing task, which suggests that if the gender effect is in fact real (which, as noted, is in question given the small number of men in each study and the unequal cells), there is a consistent and plausible explanation for it. It appears men are simply less willing to report they were distressed by the extermination task via the self-report questionnaires. However, once it comes time to wash their hands and thereby “cleanse” themselves of what they have done, their behavior matches that of the women, and they exhibit comparable levels of distress. Thus, the reported “gender effect” appears to be more of a “masculinity effect” such that the male participants in these studies are reluctant to admit their distress, yet cannot mask it when it comes to a more implicit (and perhaps more sensitive) behavioral assessment. It is also interesting to note that all of the researchers in all four studies were women. Perhaps this may have factored into the findings of the present studies, as previous research (e.g., Webber et al., 2013) used both male and female researchers and did not consistently find comparable effects with gender.

In sum, the results of Study 4 build on those of Study 2, and provide further support for the notion that self-affirmation may backfire under some circumstances, as suggested by previous research (e.g., Blanton et al., 1997). Similar to affirming on a domain of the self-concept that one has recently violated, killing does indeed appear to be a unique type of self-threat that may not be easily undone via typical self-affirmation processes.

## General Discussion

Although self-affirmation restores self-integrity in the face of self-threats (e.g., Steele & Liu, 1983), the current research was designed to examine the impact of self-affirmation on a unique type of self-threat: killing. Consistent with more recent research on self-affirmation theory (e.g., Aronson et al., 1995; Blanton et al., 1997), I believed that because killing is an act that represents a very clear departure from moral standards, self-affirmation would backfire in this instance, as one's behavior departs markedly from one's self-views. This attitude-behavior mismatch should create a self-discrepancy effect (rather than a self-affirmation effect), thereby leading to heightened levels of dissonance and distress. Thus, self-affirmation should increase the discrepancy between killing and the belief that killing is wrong (i.e., the discrepancy between one's beliefs and one's behavior), resulting in high distress and dissonance.

Across a series of four studies, the results supported this hypothesis, showing that affirming the self either failed to reduce guilt and distress after killing, or exacerbated this distress. Importantly, I obtained findings in support of the hypothesis on several dependent measures, including both self-report guilt and shame as well as a subtle behavioral measure of guilt. These findings show self-affirmation led to higher levels of both explicit (self-report) and implicit (behavioral) guilt and distress, providing converging evidence and lending more credibility to the conclusions that can be drawn. Although I did not obtain significant findings for every index of distress in every study, the pattern of means was always in the hypothesized direction, painting a consistent picture of how self-affirmation affects the distress caused by killing. Moreover, the hand washing measure of behavioral guilt generally provided the strongest and most consistent evidence, indicating implicit measures may be better suited for capturing participants' true levels of distress, as opposed to the more direct, face valid self-report

measures. Perhaps what this suggests is implicit behavioral measures may be less affected by conscious, self-presentational concerns. The “gender effect” obtained in several studies implies this may be the case, especially for male participants.

### **Addressing Inconsistent Findings across Studies**

Although the present studies generally supported the hypothesis, it is important to note that some of the studies provided stronger evidence than others did. For example, Study 1 showed a different pattern of results that contrasted the hypothesis (such that self-affirmed participants reported less guilt and distress than non-affirmed participants did). At first glance, it may be puzzling that the opposite effect was obtained in Study 1, however, after considering the changes that were made between Study 1 and Study 2, there is a plausible explanation for the observed differences. Recall that in Study 1, participants were free to choose the number of bugs they felt they needed to kill in order to get into the head of an exterminator. This allowed participants to choose the point at which they were no longer comfortable or able to accept what they were doing in the extermination task. Participants could kill up to the point they were no longer able to justify their actions (and feel as though they had fulfilled their obligation as a research participant). Although it may seem that self-affirmation under these conditions might confirm to the participant they were still a good and moral person and they had taken the moral high road by killing as few bugs as possible to satisfy the experiment, due to the methodological issues discussed, it may be erroneous to draw this conclusion. It is possible participants were not distressed if they killed only one or no bugs. Thus, the procedure for Study 2 was altered so all participants killed a specified (and large) number of bugs. This difference between Study 1 and Study 2 appears to be crucial and may account for the observed differences in results.

In Study 2 then, participants can no longer justify their behavior because they have crossed a line for what is acceptable. Thus, in Study 2, when participants cannot justify the large number of bugs they killed, engaging in self-affirmation and reflecting on personally important values (that may emphasize perceiving themselves as good and moral people) is more likely to arouse a great deal of dissonance and distress. On the one hand, participants may want to believe they are decent people with admirable qualities, but on the other hand, they just killed a large number of bugs for no good reason. Reflecting on a valued attribute of oneself appears to backfire and has the opposite effect in this instance, making one feel worse about oneself. It likely highlights the inconsistencies between how one perceives oneself and the reality of the way one just behaved, thereby exacerbating feelings of guilt and distress. The more recent research on self-affirmation theory (discussed previously) seems to provide support for this assessment (e.g., Blanton et al., 1997). Moreover, the results of Study 4 generally support this conclusion, providing a replication of Study 2.

One more finding of interest was that the timing of the self-affirmation manipulation was potentially important, as self-affirmation did not appear to help or harm participants when they self-affirmed after killing, rather than before (Study 3). Taken together, the results of the present studies provide compelling evidence that killing is a unique type of self-threat that does not appear to be easily undone by self-affirmation procedures.

### **Limitations**

**Gender.** As outlined in the Discussion for each study, several limitations need to be addressed. First, there was a consistent gender effect across the studies, indicating men and women may have responded differently. This effect was only present on the self-report indices, but did not appear on the behavioral hand washing measure. This pattern suggests men were

unwilling, for whatever reason, to report their distress (perhaps as a desire to avoid appearing “weak”), yet their hand washing behavior indicated they were still feeling quite guilty. Although this “masculinity effect” is completely reasonable given the pattern of results, it is also worth noting that the gender effect became markedly weaker across the studies, as I obtained larger samples (and more evenly distributed cells for men and women). The gender effect is therefore perhaps more convincingly attributed to an artifact of unequal samples across the genders.

**Exclusions.** Another limitation involves the unexpectedly high rates of non-compliance and attrition, particularly in Studies 3 and 4. It seemed that adding the control condition in Study 3 confused many of the participants, leading to a misunderstanding that they were killing or somehow harming the bugs, despite the fact that I provided them with explicit instructions to the contrary. In a similar vein, in Study 4, a great deal of participants opted-out of the study upon learning about the bug-killing paradigm. Given the nature of the paradigm (i.e., bug extermination), one might expect more attrition in this type of study. However, the number of exclusions might still warrant caution. I attempted to account for this limitation by simply collecting more data. The goal for each study was to obtain at least 20 participants per condition, which I met. Even though there were higher rates of attrition than initially anticipated, I kept track of the number of participants who were opting out of the study, as well as those expressing suspicion, and attempted to ensure I would ultimately end up with enough participants across conditions.

**No kill control conditions.** Although I observed the predicted differences between affirmation and no affirmation, caution is still urged when interpreting the results of the (no kill) control conditions. In Study 4, distress in the no kill control condition did not differ much from the no affirmation condition on the behavioral hand washing measure (though it did differ on the

self-report measures), which was unexpected. Although this pattern is puzzling, I offered the explanation that based on the high rate of attrition in this study, I may have been left with a sample of participants who were less likely to be distressed by the procedure because those who would have been more distressed chose not to complete the study. Thus, participants in the no affirmation condition may have been more comfortable killing the bugs and therefore exhibited less distress than in previous studies. In Study 3, the affirmation and no affirmation conditions did not differ from each other, but did differ from the no kill condition. However, there were too many unanticipated problems with the control condition in this study to draw firm conclusions.

I therefore consider the results from Studies 2 and 4 to be more conclusive on several counts. First, as noted, there were too many methodological issues in Studies 1 and 3. Specifically in Study 1, dissonance was perhaps not fully aroused due to participants having a choice in how many/whether to kill, whereas participants in Study 3 did not seem to understand they were not harming the bugs in the no kill condition. Additionally, in Study 3, the timing of the manipulation came after killing, which may have affected the results (I will address this idea later in the paper). Second, the sample sizes for Studies 2 and 4 were acceptable (around 30 participants per condition, despite withdrawals/exclusions). I therefore have greater confidence in the findings of these studies, as they demonstrate a similar pattern of results.

**Generalizability of findings.** One final limitation relates to the generalizability of the findings using the bug-killing paradigm in a laboratory context to the experience of killing in the real world. These studies were conducted to examine the conditions under which self-affirmation may exacerbate guilt and distress after killing, and the hope is that the present findings will contribute to future discussions and further research on this topic (e.g., discovering protective factors that may enable one to cope with killing). However, it is important to stress

that the bug-killing paradigm may not generalize to killing in real-world contexts. Although the bug-killing paradigm is very believable and I discussed several similarities to “real-world” killing previously, it would be misguided to suggest that inducing participants to kill bugs in a study is interchangeable to killing another human. The two acts are clearly very different, not the least of which because of the vast difference in consequences for the perpetrators and victims of the killing.

In particular, the present studies utilized the killing of targets that clearly differ from humans (although perhaps it is worth noting that genocide is often enacted against individuals who are often described as insects, cockroaches, or other vermin by the perpetrators; Martens et al., 2007). Indeed, I hardly need to mention that most people perceive bugs to be extremely different from humans in a number of important ways. For example, killing bugs is socially appropriate in many cultures, and it is reasonable to suspect that most participants had therefore killed bugs prior to participating. Thus, it is understood that completing the bug-killing paradigm in an experiment will not yield the same level of psychological distress (i.e., guilt and shame) that taking the life of another human being would. Additionally, the method of extermination differs greatly from the mode of killing typically used for both humans and insects in other contexts. I therefore urge caution when interpreting the current findings and applying them to killing humans in conflict situations.

Although using this paradigm limits the generalizability of applying the findings to what one may experience when killing another human (e.g., for soldiers at war), there are still logical reasons for employing this paradigm and it does indeed seem to serve a useful purpose. As discussed previously, this paradigm enabled the study of the psychological impact of direct killing in an experimental context, thereby allowing for the control of extraneous factors and the

manipulation of variables. These points alone make the bug-killing paradigm invaluable. In fact, research examining the differences between merely observing killing and actually carrying out the act oneself (i.e., direct killing) has shown that there are important differences between these two behaviors in terms of their psychological impact (Maguen et al., 2010). Thus, the bug-killing paradigm allowed for the investigation of factors (i.e., self-affirmation) that affect the psychological distress after direct killing, which would likely be impossible otherwise.

Additionally, it is worth noting that the results of the present studies (as well as those of previous research, e.g., Webber et al., 2013), shows that people still feel bad about killing the bugs, displaying higher guilt and distress after having done so. Moreover, in most instances, killing any living thing is generally morally proscribed. Even for bugs, killing is usually only deemed socially acceptable when it occurs for legitimate reasons (e.g., when a spider is intruding into one's home, or if insects are posing a sanitation threat, or destroying one's flowers or garden, etc.; cf. Archer & Gartner, 1992). However, when the bug poses no foreseeable or justifiable threat, killing is instead deemed immoral, perhaps even condemnable (e.g., the individual who takes great pleasure in torturing ants using a magnifying glass is viewed as potentially displaying signs of a mental disorder; e.g., see DSM-V definition of conduct disorder; American Psychiatric Association, 2013).

Taken together, it is plausible that the bug-killing paradigm may be generating insight into the psychological consequences of killing that parallel the distress experienced by those who kill other humans, albeit to a much lesser extent (Webber, 2015). Indeed, it can be argued that if the threat posed by killing bugs elicits moral emotions (e.g., exacerbating guilt and distress), then it follows that killing a human being would lead to an even higher level of distress (Martens et al., 2007). In other words, most people would experience far greater difficulty taking the life of

a human than taking the life of a few insects, and, as a result, feelings of distress are likely to be more intense compared to in the bug-killing paradigm used in the present studies. Perhaps the safest conclusion to draw is that the findings of the present studies may serve as a baseline for identifying the moral emotions elicited by killing, pointing to what one might experience when killing a human, but to a much lesser degree.

### **Framing the Findings within the Self-Affirmation and Distress Literatures**

The discussion up to this point has suggested the present research may contain certain key features that might provide a unique contribution to the self-affirmation and distress literatures. I further detail some of these aspects in the following section.

**More than self-discrepancy?** As noted, the results of the present studies indicate that self-affirmation can backfire and cause heightened levels of guilt and distress after killing, as it likely calls attention to one's morally discrepant behavior. In general, self-affirmation serves to reinforce the view that one is a good and moral person by focusing on one's positive attributes, thereby restoring one's self-integrity and the self-concept (e.g., Steele & Liu, 1983). However, killing flies in the face of this self-view, as it is an irrefutably immoral act, calling one's "goodness" into question. The present studies therefore provide support for the ideas proposed by previous researchers (e.g., Aronson et al., 1995; Blanton et al., 1997), arguing one's level of psychological distress (e.g., dissonance) will be affected by the type of affirmation one uses to restore a positive self-view. According to this research (Aronson et al., 1995; Blanton et al., 1997), if one affirms on the same aspect of the self that was threatened, this may create a self-discrepancy effect by highlighting the violated standard. This type of affirmation exacerbates negative self-appraisals by focusing on the discrepancy between one's beliefs and behavior, increasing negative emotions such as guilt and shame. A markedly different result takes place,

however, when one engages in self-affirmation in a domain not related to the self-threat. When one affirms under these circumstances, a self-affirmation effect is instead obtained, which, as noted, serves to bolster one's perception of self, restoring one's self-concept (e.g., Steele & Liu, 1983).

Building on the foundation provided by Aronson and colleagues (1995; Blanton et al., 1997), the results of the present studies indicate that perhaps one further option can be proposed. Because killing is a unique domain in which to examine the effects of self-affirmation, it seems reasonable to suggest the type of threat posed by this act will likely affect the entirety of one's self-concept, calling one's beliefs about oneself into question. Thus, more so than a mere self-discrepancy effect that arises from relatively minor inconsistencies between one's behavior and beliefs, the type of threat caused by killing is likely to be much more global, damaging one's view of the self as decent, good, and moral. As most people are likely to place a high value on these fundamental characteristics, the whole perception of self comes under threat by an act that violates these attributes at their core. Self-affirmation restores the self-concept by bolstering the moral adequacy of the self, but self-affirmation to deal with the distress of killing serves to remind one of the violated standards of morality. Under these circumstances, a heightened self-discrepancy effect will occur, such that self-affirmation will fail to alleviate dissonance, and instead will only serve to exacerbate the psychological distress (e.g., shame and guilt), rather than provide relief from the discomfort. The findings of the present studies may take the research initiated by previous researchers (Aronson et al., 1995; Blanton et al., 1997) a step further, contributing to the literature suggesting our understanding of the protective nature of self-affirmation may not be as clear-cut as originally believed.

**Timing of self-affirmation.** Another finding worth noting is the results of Study 3 suggest the timing of the self-affirmation intervention may be important in terms of whether affirmations will increase psychological distress after killing, as participants in both the affirmation and no affirmation condition displayed similar levels of distress compared to those who did not kill.

In general, current articulations of self-affirmation theory seem to suggest that self-affirmation restores a positive self-image in the face of self-threats, suggesting it should occur after encountering a threat. However, the bulk of the research tends to provide affirmation *prior* to the threat (e.g., self-affirmation is generally manipulated prior to completing a counterattitudinal essay). Thus, it is unclear whether it is better to place self-affirmation interventions before or after the self-threat. Indeed, this issue of whether or not the timing of the affirmation matters has been the subject of previous research, with mixed results. For example, McQueen and Klein (2006) conducted a systematic review of the research on self-affirmation manipulations and concluded that, of the studies reviewed, many of them did indeed vary the timing of self-affirmation to be presented both before and after the threat, and many obtained “positive effects” (McQueen & Klein, 2006, p. 302), suggesting the presentation order is not important. However, these researchers do concede that there are perhaps certain moderators that may influence whether or not the timing of the affirmation *will* matter.

In a later series of studies then, other researchers challenged this conclusion (that timing is not important), and offered the hypothesis that the crucial moderating factors surround the *response* to the threat, rather than the threat in and of itself. In other words, Critcher et al., (2010) argued that when post-threat affirmations are effective for reducing defensiveness, it is because participants had not yet engaged in defensive processes. Meaning, that in these cases,

the self-affirmation intervention is presented before participants have a chance to address the threat. Indeed, they argue “post-threat affirmations are often successful only because the defensive processing measured in affirmation studies does not start until participants are presented with the defensiveness measure itself” (Critcher et al., 2010, p. 3). Results of four studies supported these claims, showing self-affirmation may only work when it is provided before the commencement of defensive responding. In other words, participants may not actually initiate defensive processing of the threat until given an opportunity to do so when presented with the measures designed to assess defensiveness.

Perhaps what this suggests for the present studies is that, once again, killing is a unique type of self-threat because it calls the entirety of the self-concept into question. Although other self-threats are relatively subtle, killing is a very powerful and direct type of self-threat that may immediately initiate defensive processes. Thus, self-affirmation will only work (which, in this instance, may mean it will backfire) when it is conducted prior to killing. However, when it is presented after the threat, it will fail to affect participants because defensive processes commence immediately after the act. The present findings therefore suggest whether or not the timing of the affirmation intervention matters may depend on the type of threat and how quickly defensive processes are initiated. However, as noted, the methodological issues encountered in this study warrant caution in interpreting this finding and perhaps future research may wish to address this further.

### **Implications**

In addition to potentially shedding light on issues that have been presented in the self-affirmation literature (e.g., whether the timing of the manipulation matters and further elaborating on the self-discrepancy effect proposed by previous researchers), the present research

suggests killing does indeed lead to psychological distress. It is important, therefore, to discuss how this distress might relate to the psychological functioning of those who kill in the real world. Indeed, research on post-traumatic stress disorder (PTSD) suggests the guilt and shame soldiers experience can have a profound impact on their psychological health (Lee et al., 2001). Indeed, PTSD is a serious and debilitating disorder. Prevalence rates for the disorder have been increasing among some populations (e.g., rates have nearly doubled among Canadian armed forces since 2002, rising from 2.8% to 5.3% among those returning from war; Grant, 2014), which is perhaps not surprising given the number of ongoing/current armed conflicts. Thus, bearing the limitations of the present studies in mind, the findings do seem to align with information obtained from the aftermath of real-world conflict situations, particularly concerning killing and the negative impact it can have on one's mental functioning.

As suggested by the quotation at the outset of this paper, killing is “prohibited and unacceptable,” yet war is a clear example of a situation where we expect people to kill, despite clear moral prescriptions against it in the broader culture. Indeed, the reason why war is so devastating for soldiers' mental functioning is it places soldiers in situations that force them to behave in a manner that counteracts personal (as well as societal) beliefs regarding what is morally right or wrong (Litz et al., 2009; Webber et al., 2013). More specifically, war requires soldiers to perform actions (i.e., killing) that are especially threatening to their sense of self, which forces them to go against their values, morals, and beliefs that have been instilled in them from a very young age which state that aggression (and killing in particular) is morally unacceptable. That is, we ask soldiers to kill other people, while, at the same time, condemning the act.

Killing may not be the only behavior that contradicts previously held beliefs; however, research suggests it is (understandably) a critical factor in whether or not the soldier later develops PTSD. Research has revealed that soldiers who had killed in the Vietnam War later displayed higher levels of PTSD symptoms compared to soldiers who did not directly experience the killing of enemy soldiers (or civilians). Moreover, whether or not the individual had killed someone was still a major predictor of the level of PTSD symptoms, even after controlling for the intensity of battle experience (MacNair, 2002). In fact, killing was the strongest predictor of mental health dysfunction in Iraq War veterans, controlling for combat exposure (Maguen et al., 2010; Webber et al., 2013).

Additionally, research shows shame and guilt are among the key emotions thought to be involved in developing and maintaining PTSD. Although fear is often implicated as being one of the primary emotions in the formation of the disorder, shame and guilt may also play a role (Lee et al., 2001). Researchers have even argued shame and guilt may impede one's ability to process the emotions of the event, which intensifies post-traumatic symptoms. In fact, a guilt-based formulation of PTSD has recently been introduced into the clinical literature, which is identified by intense feelings of remorse connected to the event in question (Lee et al., 2001).

Given the far-reaching consequences war and killing may have on one's mental health and psychological functioning, it seems important to try to understand the factors that may mitigate the distress experienced due to taking a life. As noted above, soldiers are often placed in a troubling predicament, in which they know (and are taught) killing is wrong, yet they are often required to do so anyway. This predicament may be one reason why killing is especially problematic for them, and it is therefore an imperative concern for psychologists, in both clinical practice, as well as for researchers, to find ways to counteract or limit these negative effects.

Indeed, one of the primary motivations for the current research was to try to determine if there are certain protective factors that may help individuals cope with killing.

Unfortunately, as the results of the present studies suggest, self-affirmation does not appear to be one of these mitigating factors. As discussed, self-affirmation has long been identified in the literature as an intervention that may serve a buffering or protective function against threats to the self by bolstering self-integrity and nurturing a positive view of oneself. However, the current research supports the view that we cannot apply this buffering effect to killing, which indeed appears to be a unique type of self-threat. One practical implication of this finding is that individuals who have directly killed (e.g., soldiers at war) may be at further risk if they engage in some form of self-affirmation. That is, according to the present studies, focusing on one's positive or moral attributes may backfire and fail to provide protection from the self-threat posed by killing. Rather, the act of affirming the self will serve only to widen the attitude-behavior gap, aggravating one's feelings of guilt and shame. Importantly, this may actually foster PTSD symptoms. Indeed, as proposed in this paper, killing may violate the entirety of the self-concept, and the aspect of self being affirmed will therefore likely be of no consequence.

Based on the current findings then, perhaps self-affirmation (similar to social invalidation) may disrupt one's efforts to disengage from one's morally injurious behavior (e.g., actions undertaken by soldiers during war). This impaired ability for moral disengagement may increase one's potential for developing PTSD and lingering trauma (Webber, 2015), by making the initial act of killing more distressing because it so deeply threatens one's self-concept. Conversely, if individuals are able to disengage (or distance themselves from their behavior), they may have a chance of maintaining psychological health. This seems to be possible only when one does not engage in self-affirmation. Perhaps in this instance, one is able to convince

oneself that one's actions remain isolated in a specific place and time, in which the situation demanded killing as a legitimate means to achieve an important and desired goal (Webber, 2015). However, this may not be possible if one reflects on one's core values, as this type of self-focus only increases one's distress by increasing one's level of self-awareness and connection to the immoral behavior just performed.

Thus, the answer for potential methods for enhancing coping after killing appears to lie outside of the self. Merely affirming the self on cherished values is not enough to undo what the soldier has done. Trying to convince oneself that one is still a good and moral person in the face of knowing one has engaged in atrocities during war will not be enough to counteract one's nagging guilt. Instead, one may need reassurance from important others to convince him or her that he or she is still a person of worth and value (e.g., Webber et al., 2013). One's social support network needs to provide validation that although the individual may have done bad things (during the war), that does not make him or her a bad person. Providing this halo of warmth and support may indeed allow one to morally disengage and thereby preserve one's mental health.

## **Conclusion**

To conclude, the present and other recent research (e.g., Aronson et al., 1995; Blanton et al., 1997) suggests self-affirmation can actually produce a self-discrepancy effect, causing an increase in the moral emotions of guilt and shame. More specifically, the present analyses provided support for the hypothesis that attempting to restore self-integrity via self-affirmation processes can backfire when trying to ameliorate the distress of killing. In terms of the theoretical contributions, these results contribute to the growing list of studies on the unique effects of killing. Moreover, this research lends support to previous theorizing on the conditions

under which self-focus, or self-affirmation, can fail, and even lead to more dissonance and distress (Aronson et al., 1995; Blanton et al., 1997). Taken together, these notions suggest self-affirmation may not be the panacea for self-threats that the lion's share of work on self-affirmation would suggest.

## References

- American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders* (5<sup>th</sup> ed.). Arlington, VA: American Psychiatric Publishing.
- Archer, J. (2004). Sex differences in aggression in real-world settings: A meta-analytic review. *Review of General Psychology, 8*, 291-322. doi: 10.1037/1089-2680.8.4.291
- Archer, D., & Gartner, R. (1992). Peacetime casualties: The effects of war on the violent behavior of noncombatants. In E. Aronson (Ed.), *Readings about the social animal* (pp. 327-338). New York, NY: W. H. Freeman.
- Aronson, E. (1968). Dissonance theory: Progress and problems. In R. Abelson, E. Aronson, W. McGuire, T. Newcomb, M. Rosenberg, & P. Tannenbaum (Eds.), *Theories of cognitive consistency: A sourcebook* (pp. 5-27). Chicago: Rand McNally.
- Aronson, E. (1969). The theory of cognitive dissonance: A current perspective. *Advances in Experimental Social Psychology, 4*, 1-34.
- Aronson, J., Blanton, H., & Cooper, J. (1995). From dissonance to disidentification: Selectivity in the self-affirmation process. *Journal of Personality and Social Psychology, 68*, 986-996.
- Bandura, A. (1990). Mechanisms of moral disengagement in terrorism. In W. Reich (Ed.), *Origins of terrorism: Psychologies, ideologies, states of mind* (pp. 161-191). Cambridge: Cambridge University Press.
- Baum, D. (2004). The price of valor: We train our soldiers to kill for us. Afterward, they're on their own. *The New Yorker, 19*, 1-7.

- Blanton, H., Cooper, J., & Aronson, J. (1997). When bad things happen to good feedback: Exacerbating the need for self-justification with self-affirmations. *Personality and Social Psychology Bulletin*, *23*, 684-692.
- Brehm, J. & Cohen, A. (1962). *Explorations in cognitive dissonance*. New York: Wiley
- Card, N. A., Stucky, B. D., Sawalani, G. M., & Little, T. D. (2008). Direct and indirect aggression during childhood and adolescence: A meta-analytic review of gender differences, intercorrelations, and relations to maladjustment. *Child Development*, *79*, 1185-1229.
- Cohen, G. L., Aronson, J., & Steele, C. M. (2000). When beliefs yield to evidence: Reducing biased evaluation by affirming the self. *Personality and Social Psychology Bulletin*, *26*, 1151-1164. doi: 10.1177/01461672002611011
- Cohen, G. L., & Sherman, D. K. (2007). Self-affirmation theory. *Baumeister R, Vohs K, Encyclopedia of Social Psychology*. Thousand Oakes: Sage Publications.
- Cooper, J. (1971). Personal responsibility and dissonance: The role of foreseen consequences. *Journal of Personality and Social Psychology*, *18*, 354-363.
- Cooper, J., & Brehm, J. W. (1971). Prechoice awareness of relative deprivation as a determinant of cognitive dissonance. *Journal of Experimental Social Psychology*, *7*, 571-581.
- Cooper, J., & Fazio, R. H. (1984). A new look at dissonance theory. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 17, pp. 229-266). New York: Academic Press.
- Cooper, J., & Worchel, S. (1970). Role of undesired consequences in arousing cognitive dissonance. *Journal of Personality and Social Psychology*, *16*, 199-206.

- Cougle, J. R., Goetz, A. R., Hawkins, K. A., & Fitch, K. E. (2012). Guilt and compulsive washing: Experimental tests of interrelationships. *Cognitive Therapy and Research, 36*, 358-366. doi: 10.1007/s10608-011-9359-x
- Critcher, C. R., Dunning, D., & Armor, D. A. (2010). When self-affirmations reduce defensiveness: Timing is key. *Personality and Social Psychology Bulletin, 36*, 947-959. doi: 10.1177/0146167210369557
- Crocker, J., Niiya, Y., & Mischkowski, D. (2008). Why does writing about important values reduce defensiveness? Self-Affirmation and the role of positive other-directed feelings. *Psychological Science, 19*, 740-747. doi: 10.1111/j.1467-9280.2008.02150.x
- Duval, S., & Wicklund, R. A. (1972). *A theory of objective self-awareness*. New York: Academic Press
- Federoff, N. A., & Harvey, J. H. (1976). Focus of attention, self-esteem, and the attribution of causality. *Journal of Research in Personality, 10*, 336-345.
- Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford, CA: Stanford University Press.
- Fleiss, J. L. (1981). *Statistical methods for rates and proportions* (2nd ed.). New York, NY: Wiley.
- Gibbons, F. X. (1987). Mild depression and self-disclosure intimacy: Self and others' perceptions. *Cognitive Therapy and Research, 11*, 361-380.
- Gibbons, F. X. (1990). Self-attention and behavior: A review and update. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 23, pp. 249-295). New York: Academic Press.
- Grant, K. (2014, August 11). Post-traumatic stress disorder doubles among Canadian Forces. *The Globe and Mail*. Retrieved from <http://www.theglobeandmail.com/>

- Greenberg, J., & Musham, C. (1981). Avoiding and seeking self-focused attention. *Journal of Research in Psychology, 15*, 191-200.
- Hagen, J. (1994). *Crime and Disrepute*. Thousand Oaks, CA: Pine Forge Press.
- Harber, K. (1995). *Sources of validation scale*. Unpublished scale.
- Harris, P. R., & Napper, L. (2005). Self-affirmation and the biased processing of threatening health-risk information. *Personality and Social Psychology Bulletin, 31*, 1250–1263. doi: 10.1177/0146167205274694
- Higgins, E. T. (1987). Self-discrepancy: A theory relating self and affect. *Psychological Review, 94*, 319-340.
- Keough, K. A., Garcia, J., & Steele, C. M. (1998). Reducing stress and illness by affirming the self. Unpublished manuscript, Stanford University.
- Koole, S. L., Smeets, K., van Knippenberg, A., & Dijksterhuis, A. (1999). The cessation of rumination through self-affirmation. *Journal of Personality and Social Psychology, 77*, 111–125.
- Kubany, E. S., Haynes, S. N., Abueg, F. R., Manke, F. P., Brennan, J. M., & Stahura, C. (1996). Development and validation of the trauma-related guilt inventory (TRGI). *Psychological Assessment, 8*, 428-444. doi: 10.1037/1040-3590.8.4.428
- Lee, D. A., Scragg, P., & Turner, S. (2001). The role of shame and guilt in traumatic events: A clinical model of shame-based and guilt-based PTSD. *British Journal of Medical Psychology, 74*, 451-466. doi: 10.1348/000711201161109
- Lewis, H. B. (1971). *Shame and guilt in neurosis*. New York: International Universities Press

- Litz, B. T., Stein, N., Delaney, E., Lebowitz, L., Nash, W. P., Silva, C., & Maguen, S. (2009). Moral injury and moral repair in war veterans: A preliminary model and intervention strategy. *Clinical Psychology Review, 29*, 695-706. doi: 10.1016/j.cpr.2009.07.003
- Liu, T. J., & Steele, C. M. (1986). Attributional analysis as self-affirmation. *Journal of Personality and Social Psychology, 51*, 531-540.
- MacNair, R. M. (2002). Perpetration-induced traumatic stress in combat veterans. *Peace and Conflict: Journal of Peace Psychology, 8*, 63-72.
- Maguen, S., Lucenko, B. A., Reger, M. A., Gahm, G. A., Litz, B. T., Seal, K... & Marmar, C. R. (2010). The impact of reported direct and indirect killing on mental health symptoms in Iraq war veterans. *Journal of Traumatic Stress, 23*, 86-90. doi: 10.1002/jts.20434
- Marschall, D., Santfner, J., & Tangney, J. P. (1994). *The state shame and guilt scale*. Fairfax, VA: George Mason University.
- Martens, A., Johns, M., Greenberg, J., & Schimel, J. (2006). Combating stereotype threat: The effect of self-affirmation on women's intellectual performance. *Journal of Experimental Social Psychology, 42*, 236-243. doi:10.1016/j.jesp.2005.04.010
- Martens, A., & Kosloff, S. (2012). Evidence that killing escalates within-subjects in a bug-killing paradigm. *Aggressive Behavior, 38*, 170-174. doi: 10.1002/ab.21412
- Martens, A., Kosloff, S., Greenberg, J., Landau, M. J., & Schmader, T. (2007). Killing begets killing: Evidence from a bug-killing paradigm that initial killing fuels subsequent killing. *Personality and Social Psychology Bulletin, 33*, 1251-1264. doi: 10.1177/0146167207303020

- Martens, A., Kosloff, S., & Jackson, L. E. (2010). Evidence that initial obedient killing fuels subsequent volitional killing beyond effects of practice. *Social Psychological and Personality Science, 1*, 268-273. doi: 10.1177/1948550609359813
- McQueen, A., & Klein, W. M. (2006). Experimental manipulations of self-affirmation: A systematic review. *Self and Identity, 5*, 289-354. doi: 10.1080/15298860600805325
- Nel, E., Helmreich, R., & Aronson, E. (1969). Opinion change in the advocate as a function of the persuasibility of his audience: A clarification of the meaning of dissonance. *Journal of Personality and Social Psychology, 12*, 117-124.
- Nelissen, R. M. A., & Zeelenberg, M. (2009). When guilt evokes self-punishment: Evidence for the existence of a dooby effect. *Emotion, 9*, 118-122. doi: 10.1037/a0014540
- Pallant, J. (2007). *SPSS survival manual*. New York: McGraw-Hill.
- Reed, M. B., & Aspinwall, L. G. (1998). Self-affirmation reduces biased processing of health-risk information. *Motivation and Emotion, 22*, 99-132.
- Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.
- Scheier, M. F., & Carver, C. S. (1977). Self-focused attention and the experience of emotion: Attraction, repulsion, elation, and depression. *Journal of Personality and Social Psychology, 35*, 625-636.
- Scher, S. J., & Cooper, J. (1989). Motivational basis of dissonance: The singular role of behavioral consequences. *Journal of Personality and Social Psychology, 56*, 899-906.
- Schmeichel, B. J., & Martens, A. (2005). Self-affirmation and mortality salience: Affirming values reduces worldview defense and death-thought accessibility. *Personality and Social Psychology Bulletin, 31*, 658-667. doi: 10.1177/0146167204271567

- Schmeichel, B. J., & Vohs, K. (2009). Self-affirmation and self-control: Affirming core values counteracts ego depletion. *Journal of Personality and Social Psychology, 96*, 770-782. doi: 10.1037/a0014635
- Sherman, D. K., & Cohen, G. L. (2006). The psychology of self-defense: Self-affirmation theory. *Advances in Experimental Social Psychology, 38*, 183-242.
- Shrira, I., & Martin, L. L. (2005). Stereotyping, self-affirmation, and the cerebral hemispheres. *Personality and Social Psychology Bulletin, 31*, 846-856. doi: 10.1177/0146167204272232
- Steele, C. M. (1975). Name calling and compliance. *Journal of Personality and Social Psychology, 31*, 361-369.
- Steele, C. M. (1988). The psychology of self-affirmation: Sustaining the integrity of the self. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 21, pp. 261-302). San Diego, CA: Academic Press.
- Steele, C. M., & Liu, T. J. (1983). Dissonance processes as self-affirmation. *Journal of Personality and Social Psychology, 45*, 5-19.
- Strachan, E., Schimel, J., Arndt, J., Williams, T., Solomon, S., Pyszczynski, T., & Greenberg, J. (2007). Terror mismanagement: Evidence that mortality salience exacerbates phobic and compulsive behaviors. *Personality and Social Psychology Bulletin, 33*, 1137-1151. doi: 10.1177/0146167207303018
- Tangney, J. P., Niedenthal, P. M., Covert, M. V., & Barlow, D. H. (1998). Are shame and guilt related to distinct self-discrepancies? A test of Higgins's (1987) hypotheses. *Journal of Personality and Social Psychology, 75*, 256-268.

- Tesser, A., Martin, L. L., & Cornell, D. P. (1996). On the substitutability of self-protective mechanisms. In J. A. Bargh (Ed.), *The psychology of action: Linking cognition and motivation to behavior* (pp. 48–68). New York: Guilford Press.
- Thomaes, S., Bushman, B. J., de Castro, B. O., Cohen, G. L., & Denissen, J. J. A. (2009). Reducing narcissistic aggression by buttressing self-esteem. *Psychological Science, 20*, 1536-1542. doi: 10.1111/j.1467-9280.2009.02478.x
- Warner, R. M. (2008). *Applied statistics*. California: Sage Publications
- Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54*, 1063-1070.
- Webber, D., Schimel, J., Martens, A., Hayes, J., & Faucher, E. H. (2013). Using a bug-killing paradigm to understand how social validation and invalidation affect the distress of killing. *Personality and Social Psychology Bulletin, 39*, 470-481. doi: 10.1177/0146167213477891
- Webber, D. (2015). *Preparing people to kill: A study of moral disengagement and threat perception within a bug-killing paradigm* (Unpublished doctoral dissertation). University of Alberta, Edmonton, Alberta, Canada.
- Wicklund, R. A., & Brehm, J. W. (1976). *Perspectives on cognitive dissonance*. Hillsdale, NJ: Erlbaum.
- Williams, T. J., Schimel, J., Hayes, J., & Usta, M. (2014). Following and resisting body image ideals in advertising: The moderating role of extrinsic contingency focus. *Self and Identity, 13*, 398-418. doi: 10.1080/15298868.2013.836133