When Self-Affirmations Reduce Defensiveness: Timing Is Key

Personality and Social Psychology Bulletin
36(7) 947–959

DOI: 10.1177/0146167210369557
http://pspb.sagepub.com

Clayton R. Critcher[1], David Dunning[2], and David A. Armor[3]

Abstract

Research on self-affirmation has shown that simple reminders of self-integrity reduce people's tendency to respond defensively to threat. Recent research has suggested it is irrelevant whether the self-affirmation exercise takes place before or after the threat or the individual's defensive response to it, supposedly because the meaning of threats is continuously reprocessed. However, four experiments revealed that affirmations may be effective only when introduced *prior* to the initiation of a defensive response. Affirmations introduced before threatening feedback reduced defensive responding; affirming after a threat was effective in reducing defensiveness only if the defensive conclusion had yet to be reached. Even though threats may activate a defensive motivation, the authors' results suggest that defensive responses may not be spontaneous and may be prompted only when suggested by the dependent measures themselves. This explains why some affirmations positioned after threats are effective in reducing defensiveness. Implications for self-affirmation theory are discussed.

Keywords

self-affirmation, defensiveness, threat, self-integrity, dissonance

Received March 5, 2009; revision accepted December 18, 2009

Whether from bosses, spouses, or some other critical source, people occasionally receive threatening feedback about their competence and character. Researchers have identified an eclectic array of defensive strategies that people use to dampen the impact of unfavorable information on their self-integrity, thereby allowing people to maintain unrealistically positive illusions about themselves and their place in the world (Alicke, Klotz, Breitenbecher, Yurak, & Vredenburg, 1995; Dunning, 2003; Taylor & Brown, 1988). For example, people engage in downward social comparisons (Spencer, Fein, & Lomore, 2001; Taylor & Lobel, 1989), view their own successes as unique and shortcomings as commonplace (Campbell, 1986; Marks, 1984), and dissociate their own specific deficits from broader implications for the self (Beauregard & Dunning, 2001; Wentura & Greve, 2003). Although these defensive strategies might help one brush off a crush's rejection, a less-than-flattering teaching evaluation, or a personal insult heard through the grapevine, these same defensive strategies may also prevent people from attaining accurate assessments of their own strengths and shortcomings (Dunning, 2005; Dunning, Heath, & Suls, 2004; Sedikides, Green, & Pinter, 2004). It would at times be useful to be able to "switch off" one's defensive shield so that threatening information and ideas could be more objectively considered (Critcher, Helzer, & Dunning, in press; Radcliffe & Klein, 2002).

The psychology of self-affirmation suggests that even simple reminders of self-worth may be sufficient to flip this switch—to reduce the normal tendency to respond to threat defensively—so that people can incorporate useful but potentially unflattering information about themselves (also see Trope & Pomerantz, 1998). Self-affirmation theory (Sherman & Cohen, 2006; Steele, 1988) proposes that any strategy that restores the integrity of the self after a psychic assault should alleviate the impact of the new threat and thus eliminate the need to respond defensively. In other words, threats to the self need not be dealt with at the site of the psychic wound but can be healed more indirectly by calling to mind valued aspects of one's identity in some other life domain, even though these identity considerations bear no relation to the source of the threat.

Consistent with this argument, self-affirmations have been shown to reduce a number of defensive processes (for a review, see Sherman & Cohen, 2006). Self-affirmed participants provide help to others even when that other person's success is threatening (Tesser, Martin, & Cornell, 1996), more objectively and less defensively evaluate the arguments of an ideological opponent (Cohen, Aronson, & Steele,

[1]University of California, Berkeley
[2]Cornell University, Ithaca, NY, USA
[3]San Diego State University, San Diego, CA, USA

Corresponding Author:
Clayton R. Critcher, University of California, Berkeley, Haas School of Business, Department of Marketing, Berkeley, CA 94720
Email: clayton.critcher@aya.yale.edu

2000), and are more open to self-relevant information about risks to their health, such as the link between cancer and alcohol consumption (Harris & Napper, 2005; Reed & Aspinwall, 1998).

The number and diversity of these demonstrations serve both as a sign of the strength of the self-affirmation model and a possible source of concern. Self-affirmations clearly curtail defensive responding in many circumstances, but there may be reason to suspect that several recent claims of the effects of self-affirmations may be overstated and that self-affirmations may be subject to a more significant boundary condition than past research has suggested (also see Sherman et al., 2009). In particular, the defense-inhibiting power of self-affirmations may depend on a variable that, to date, has been viewed as relatively unimportant—the timing of the affirmation in relation to the defensive response.

The importance of timing in self-affirmation processes can be predicted from the convergence of two theoretical notions that underlie current understanding of the psychology of defense: the substitutability of self-esteem maintenance mechanisms and the notion that people seek to maintain and not necessarily maximize their positive self-views. From the substitutability perspective, self-affirmation is not in itself a special process but rather one tool in an arsenal of psychic tactics that people use to bolster and maintain their self-esteem, such as finding worse-off others to compare themselves against and dispelling dissonant thoughts by altering belief in those thoughts (for a review, see Tesser, 2000). Importantly, data suggest that these tactics for self-esteem maintenance are quite interchangeable: One tactic can often be used to substitute for another (Hart, Shaver, & Goldberg, 2005; Tesser, 2000).

The second key idea, implicit in the notion of the substitutability of self-maintenance processes, is that there is a natural limit to people's positivity strivings. Results from several independent research programs suggest that people do not self-enhance at all costs (Markus & Wurf, 1987) but are motivated to maintain rather than to maximize a positive self-image (Tesser, 1988) and that their "psychological immune system" kicks in only when the self experiences some degree of threat (Aspinwall, 1997; Tesser, 1988). This self-image maintenance perspective suggests that, once a positive self-image has been restored—whether through a self-affirmation or through a defensive process—there will be no further need to restore the self as there will be no threat for any subsequent affirmation to undo.

The combination of these two notions suggests two boundary conditions that may limit the impact of self-affirmations, one of which has already received support in the affirmation literature, one of which has on its surface been consistently contradicted. First, if affirmations target defensiveness in response to threat, then affirmations should affect only the responding of threatened participants. Several studies support this threat–repair stipulation: that affirmations do halt defensive responding only of those for whom a valued self-relevant domain has been threatened (e.g., Harris & Napper, 2005).

The second condition, building on the first, centers on the timing of the self-affirmation exercise. If affirmations mostly alleviate threats to the self, then affirmations should not affect those who have already engaged in an act of defensiveness—that is, on those who have already removed the threat. In short, from these two perspectives, there is no reason to expect that affirmations would have a retroactive effect on defensive responses previously made. Affirmations should reduce any remaining inclination to be defensive but should not undo products of past defensiveness. This was the key hypothesis we explored in this research: Do affirmational interventions have an impact once a person has had a chance to respond defensively? According to our perspective, they should not.

Does Timing Matter? Competing Perspectives

On close inspection, our timing hypothesis may have been implied in Steele's (1988) formulation of self-affirmation theory. In a speculative (but, we think, prescient) paragraph, Steele wrote, "Self-affirming thoughts may be an effective means of reducing thought-distorting defense mechanisms such as denial and rationalization," but these effects "may depend, at least partially, on what other thoughts about the self are salient *at the time the information is processed* [italics added]" (p. 290). This statement clearly implies that affirmations may be expected to be effective only in prospect—that affirming thoughts must be in place as one thinks through the implications of a threat to the self.

But much evidence appears to speak against the timing hypothesis. For example, in a recent meta-analysis, McQueen and Klein (2006) found that the effect of affirmations placed after the threat was equivalent to that of those placed before the threat. Data such as these have led other researchers to modify Steele's (1988) argument to justify why affirmations should be able to undo defensive responses. Cohen et al. (2000), for example, argued that affirmations could do more than prevent defensive responding, suggesting that affirmations could also actively undo defensive conclusions that had already been drawn. In differentiating the defensiveness-reducing effects of their pre-threat affirmations from the defensiveness-reducing effects of their post-threat affirmation, Cohen et al. stated, "The affirmation may reduce on-line defensiveness processing, at the time of encoding . . . or [a post-threat affirmation] may attenuate memory-based defensive processing" (p. 1162). Cohen and colleagues posited that threats to the self are continuously reprocessed and that self-affirmations enable people to reconsider and reverse defensive conclusions even after they have been drawn.

Sherman, Nelson, and Steele (2000) advanced a similar argument to explain the effectiveness of their post-threat affirmation, "given that [the threat] is continually reprocessed and reformulated until the participant is asked to report on the attitude" (p. 1056).

The reprocessing hypothesis is thus consistent with existing data but remains stubbornly inconsistent with the theoretical principles motivating our conclusion that timing matters. In contrast to the reprocessing-based claim that the timing of the self-affirmation is irrelevant (Cohen et al., 2000; Sherman et al., 2000), we hypothesize that self-affirmations will work only in prospect—that is, only if the affirmation is in place before the threat itself or before defensive response has been initiated. If the defensive reaction has already taken place, self-affirmation may reduce any remaining motivation to respond defensively, but the self-affirmation should not "undo" a defensive reaction that has already been crafted, according to the logic of substitutability and of self-image maintenance.

How, then, can this timing hypothesis be reconciled with previous research showing that post-threat affirmations can be effective in curtailing a defensive response? We propose that the critical moderator is the timing of the affirmation in relation to a threat *response* rather than the timing of that affirmation in relation to the threat itself. To date, research on the effects of post-threat affirmations may have obscured this distinction. We argue that post-threat affirmations are at times successful in reducing defensiveness not because threats are continually reprocessed but because many post-threat affirmations are actually positioned before the response to the threat commences. In particular, we suggest that 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.

Overview of Studies

Four studies examine whether the timing of self-affirmations, in relation to the initiation of defensive responding, matters. In Study 1, we examined whether self-affirmation exercises at times reduce defensive responding more effectively when they come before rather than after a threat, establishing that timing does matter. In Studies 2a and 2b, we examined the timing of affirmation exercises more carefully to see if an affirmation is effective after the threat has occurred but before the person has had a chance to respond to it. According to our analysis, post-threat affirmations can be effective if people have yet to initiate a defensive response. However, if we can prompt people to respond defensively at an earlier moment before confronting an affirmation intervention, that intervention should lose its power to reduce defensiveness. Study 3 examined the same issues in the realm of cognitive dissonance.

Study 1

Participants were asked to take one of two versions of a test purporting to measure "integrative orientation ability." We constructed a particularly difficult version of the test meant to inspire threat. Participants were randomly assigned to self-affirm just before the test (pre-threat affirmation), just after scoring their own test (post-threat affirmation), or not at all (no affirmation). We predicted that the affirmation would be effective in reducing defensiveness in the pre-threat condition, relative to the no affirmation condition, but not in the post-threat affirmation condition. If, however, timing does not matter, then both pre- and post-threat affirmations should be equally effective in leading participants to less defensively accept the implications of their performance in the hard test condition.

As a supplemental goal, we also examined the impact of self-affirmation for individuals not threatened. We included another version of the test that was more moderately demanding and was meant to be nonthreatening (easy test). We predicted that self-affirmation would have no impact on people's defensive responding in this condition, given the view that people more enthusiastically seek to maintain their self-esteem once threatened than they do to maximize their self-esteem at all times. An impact of affirmation on the threatened, but not on the nonthreatened, would support one of the premises on which our timing hypothesis was based.

Method

Participants and design. In exchange for either class credit or $5, 184 Yale University undergraduates took part in the study. The experiment used a 3 (affirmation: pre-test affirmation, post-test affirmation, or no affirmation) × 2 (test difficulty: hard or easy) factorial design.

Procedure: Test of integrative orientation ability. All participants were given a test that ostensibly evaluated people's ability "to think creatively" and "to find unusual solutions to problems." We modified the items on Mednick's (1962) Remote Associates Test to create two versions—one that was quite difficult (and thus threatening) and one that was much easier (less threatening). Each item on these 15-item tests provides participants with three words and asks the respondent to generate a fourth word that relates to each of the other three words. A representative item from the easy test is "Chocolate—Fortune—Tin"; a representative item from the hard test is "Soap—Shoe—Tissue."[1] After a 4-min testing period, participants were given correct solutions and were asked to score their own tests.

Procedure: Self-affirmation. Participants assigned to one of the self-affirmation conditions completed a self-affirmation task adapted from previous research (Cohen et al., 2000; Shira & Martin, 2005). In this task, participants were given a list of eight domains (e.g., religious fulfillment, physical

health) and asked to rank the domains in order of personal importance. Because affirming on the threatened domain can produce a backfiring defensiveness-exaggerating effect (Blanton, Cooper, Skurnik, & Aronson, 1997; Sivanthan, Molden, Galinsky, & Ku, 2008), none of the domains related to academic or intellectual achievement. Participants then wrote a paragraph about their most valued domain's importance in their lives. In place of the self-affirmation task, control participants received a list of exotic-sounding jelly beans and were asked to rank them by how tasty they believed the flavors would be.[2] Pre-test affirmation participants completed the affirmation just before taking the test. Post-test affirmation participants completed the exercise just after scoring their test but before the measures of defensiveness. To control for the delay the post-test affirmation caused between the test and the critical measures, both pre-test affirmation and control participants completed the jelly bean task after scoring their own test.

Procedure: Measures of defensiveness. Participants estimated the average score, out of 15, that they thought students at Yale University would achieve on the test. By minimizing their estimates of the average score, participants could defensively paint their own performance in a better light (Klein & Kunda, 1989). Second, we obtained an assessment of how much participants incorporated feedback into their self-concepts. To make this measure less transparent, we presented it as part of a seemingly unrelated study in which participants were asked to rate themselves and an acquaintance on a variety of traits. Within this "second study," participants were asked to rate their own creativity on a scale from 1 (*not at all*) to 9 (*completely*). As the instructions to the integrative orientation ability test had indicated that the test was a measure of creative thinking, higher self-ratings of creativity evidenced greater defensiveness.

Results

Manipulation check. Test scores were submitted to a 3 (affirmation: pre-test affirmation, no affirmation, or post-test affirmation) × 2 (test version: hard or easy) ANOVA. As expected, participants performed significantly worse on the hard test ($M = 2.8$) than on the easy test ($M = 10.9$), $F(1, 178) = 596.71$, $p < .001$. Neither the main effect of affirmation nor the interaction with test condition was significant, $F(2, 178) = 1.89$, $p > .15$, and $F(2, 178) = 1.59$, $p > .20$, respectively, suggesting that being affirmed did not affect one's performance.

Influence of threat and timing. To maximize power in detecting differences in defensiveness, we standardized each of the two measures of defensiveness, reverse scored the perceived average score, and then summed the two measures. According to the threat–repair stipulation, an ANOVA on the defensiveness index should reveal a threat (easy or hard test) by affirmation (pre-test affirmation, post-test affirmation, or no affirmation) interaction, with any defensive-reducing effects of pre-test (and possibly post-test) affirmations expected only in the threat (hard-test) conditions. As depicted in Figure 1, the effect of the affirmation manipulation did depend on whether participants took the hard or easy test, $F(2, 173) = 4.50$, $p = .01$. As predicted, the influence of affirmation was significant for those who took the hard (threatening) test, $F(2, 96) = 5.86$, $p = .004$, but not for those who took the easy one, $F(2, 77) = 1.05$, $p > .35$.

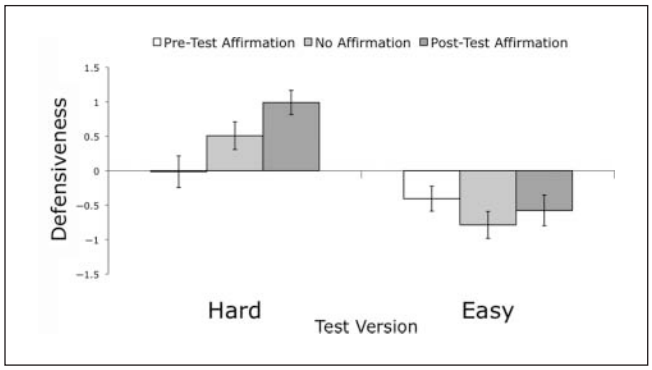

Figure 1. Defensiveness as a function of affirmation condition and test difficulty (Study 1)
Values are z score composites and thus have no absolute meaning; 0 is not "no defensiveness." Comparisons within the easy and hard test conditions are meaningful comparisons of defensiveness; comparisons between these conditions are not. Instead, between-test-condition differences reflect a mix of defensiveness and actual informational differences.

To distinguish between the timing and reprocessing hypotheses, we then placed responses in the hard test condition under closer scrutiny. We conducted two contrasts to determine whether affirmations were effective only when administered before the threat (pre-threat: –2; post-threat and control: +1) or whether affirmations were effective regardless of timing (pre-threat, post-threat: –1; control: +2). These contrasts suggested that affirmations were effective only if encountered before the threat, $t(96) = 3.02$, $p = .003$.[3] The competing contrast inspired by the reprocessing hypothesis was not significant ($t < 1$). Follow-up comparisons found that pre-threat affirmations marginally reduced defensiveness compared to the control condition, $t(96) = 1.83$, $p = .07$, but crucially were more effective in reducing defensiveness than were post-threat affirmations, $t(96) = 3.42$, $p = .001$. Post-threat affirmed participants unexpectedly displayed marginally more defensiveness than those in the control condition, $t(96) = 1.70$, $p = .09$. There was no hint of this pattern in future studies, so we hesitate to speculate on this effect.

Discussion

The results of Study 1 are consistent with the timing hypothesis. Only participants who had been affirmed prior to taking the difficult test (and thus prior to the receipt of critical

feedback) showed a reduction in defensive responding. Contrary to the notion that defensive conclusions are continually reassessed, self-affirmations did not undo defensive conclusions that, presumably, had already been formed. An act of defensiveness does not appear to be merely a temporary brace that bolsters the damaged self until a self-affirmation's calming force can allow for dispassionate reprocessing. Instead, affirmations, in restoring self-integrity, appear to obviate the need for further acts of defensiveness while leaving already formulated defensive conclusions firmly in place.

Studies 2a and 2b

Our timing hypothesis does not preclude the possibility that post-threat affirmations will, at times, be effective in preventing defensiveness. A post-threat affirmation could be effective in eliminating defensive processes if (and perhaps only if) these processes are not executed until the participant is given a chance to respond defensively. That is, the timing of the threat per se does not matter as much as the timing of the individual's response to that threat. An affirmation taking place after a threat can be effective if it precedes the individual's response to that threat. However, according to the reprocessing hypothesis, a post-threat affirmation should still be effective even if it were introduced after the individual has responded to a threat.

Past work on self-affirmation has shown that post-threat affirmations can be effective in quelling defensive responding (e.g., McQueen & Klein, 2006). It is possible that the specific defensive strategies under investigation in these past studies may not have been spontaneous. That is, people may not have initiated their defensive responding until given the post-threat questionnaire that allowed them the opportunity to be defensive. Thus, there may have been gaps in time between the threat and participants' reactions to it in which self-affirmation could still be effective. If this is the case, then previous studies using seemingly effective post-threat affirmations may not have provided evidence of reprocessing at all but may instead have demonstrated the effect of an affirmation on a defensive response that had yet to begin.

To more directly test the effects of the timing of a self-affirmation in relation to a threat response and not just in relation to the threat itself, it was necessary to experimentally disentangle participants' awareness of a defensive opportunity from the time at which they could provide a defensive response. Studies 2a and 2b sought to do just that. In Study 2a, we created conditions in which defensive reactions to threat would likely not be spontaneously initiated (giving post-threat affirmations a chance to be effective) by making opportunities to respond defensively nonobvious and by limiting the time between the experience of threat and the administration of the post-threat affirmation.

Then, in Study 2b, we not only replicated these conditions but also added a foreshadowing condition that informed participants, before completing a self-affirmation exercise, of defensive opportunities we were going to present them on a follow-up questionnaire. This foreshadowing was intended to unconfound the presentation of the defensive measures (and thus the suggestion of particular strategies for ego repair) from the elicitation of the defensive response. We hypothesized that post-threat affirmations would reduce defensiveness when these affirmations were introduced before the threat response (Study 2a) but not when possibilities for defensive processing had already been brought to the attention of participants (Study 2b). Although Study 1 used threats to intellectual abilities, Studies 2a and 2b used a threat to participants' ability to maintain and foster personal relationships (Baumeister & Leary, 1995).

Study 2a

In Study 2a, we changed the nature of the threat to better control when participants might start a defensive response to threat. In Study 1, it seemed likely that participants in the high-threat (hard-test) conditions knew they were doing poorly for the entire duration of the test-taking experience (they were not able to come up with free-response answers), and thus defensiveness may have set in early. To prevent this, in Studies 2a and 2b participants were given a test that did not offer performance-related cues, and evaluative feedback was withheld until after the test. Second, we chose indirect, nonobvious measures of defensiveness that participants were unlikely to engage in until made aware of them in the questionnaire.

For Study 2a, we predicted that affirmation, relative to a control condition, would quell defensive responding regardless of whether participants affirmed before or after receiving the threatening feedback. If this result emerged, we could test in Study 2b whether the post-threat affirmation was effective because defensive processing had yet to be engaged.

Method

Participants and design. In exchange for extra credit in psychology and human development courses, 76 Cornell University undergraduates took part in our study. Participants were randomly assigned to one of three high-threat affirmation conditions: a pre-threat affirmation (administered before the test), a post-threat affirmation (administered after both the test and feedback but before measures of defensiveness), or a no affirmation control.

Procedure: Test of interpersonal perception ability. All participants completed the 15-item version of the Interpersonal Perception Task (IPT-15; Costanzo & Archer, 1989). The IPT-15 is a video-based test that participants were told would evaluate their skill at "accurately perceiving verbal and nonverbal interpersonal cues," which was said to be a crucial skill in "fostering and maintaining interpersonal

relationships." The test comprises 15 scenes that participants watched on a computer. After each scene, the participants answered a multiple-choice question about the scene. For example, in one scene, the test taker watches a short conversation between a man and a woman sitting at a table. Based on verbal and nonverbal cues, participants must decide whether the pair are siblings or newly formed friends. Following the 20-min test, the experimenter informed participants that to protect their anonymity, test takers would score their own tests. After exchanging participants' blue pens for red pens to prevent dishonesty, the experimenter provided each test taker with an answer key. In actuality, the responses on the answer key had been randomly generated, creating falsely negative feedback for the participants. Participants rated on a 9-point scale whether their performance was *much worse* (1) or *much better* (9) than expected.

Procedure: Self-affirmation. As in the previous study, the self-affirmation manipulation asked participants to write about an important value. To avoid having participants affirm themselves on a threat-relevant domain, the value "helping others" was replaced with "academic success." Participants in the pre-threat affirmation condition were asked to complete the affirmation before reading the test instructions. Those in the post-threat affirmation condition were asked to complete the affirmation immediately after scoring the test but prior to the measures of defensiveness. Participants assigned to the no affirmation control condition were presented with the names of different candle scents (e.g., coconut smoothie, baby powder) and were asked to rank order these scents in terms of preference. Control participants, in addition to pre-threat affirmation participants, completed the candle-rating task while post-threat affirmation participants were completing the self-affirmation.

Procedure: Measures of defensiveness. We chose four measures of defensiveness that seemed likely to capture nonspontaneous processes. First, we asked participants to evaluate their own intelligence so as to give participants an opportunity to "compensate" for their poor performance by exaggerating their ability in another domain (Brown & Smart, 1991). Second, given that the test was supposed to tap their ability to foster and maintain interpersonal relationships, we asked participants to estimate how many friends they had and how many acquaintances they had. Given that participants were expected to still be smarting about their negative score on a test that was supposed to tap their ability to foster and maintain interpersonal relationships, we reasoned that asking about their personal relationships would provide them with an opportunity to reaffirm their (previously threatened) ability to form and maintain many personal relationships. Finally, we told participants that there was another test that had been matched for difficulty with the version that they had just taken. We asked them what score they thought they would receive on that test. For each measure—self-rated intelligence, number of friends, number of acquaintances, and score on a retest—higher values reflect greater defensiveness.

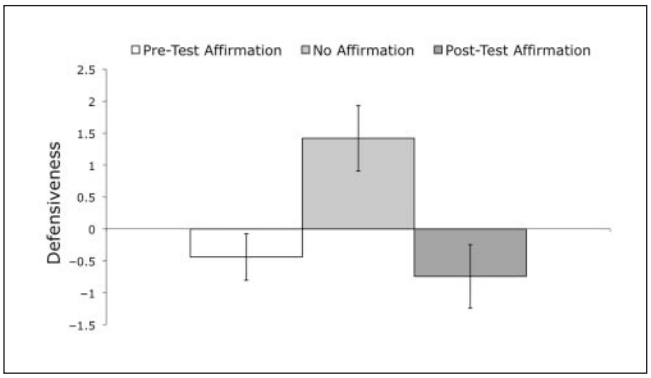

Figure 2. Defensiveness as a function of affirmation condition (Study 2a)
Values are z score composites and thus have no absolute meaning; 0 is not "no defensiveness."

Results

Perceptions of test performance. A preliminary analysis provided assurance that the random-response answer key provided falsely negative feedback to participants. Participants' responses on the rating of performance relative to their expectation showed that they tended to rate their score in the "worse than expected" side of the scale ($M = 3.9$ vs. a scale midpoint of 5.0), $t(75) = 6.63$, $p < .001$.

Measures of defensiveness. As in Study 1, we began by standardizing and summing the measures of defensiveness. The means by condition are displayed in Figure 2. We conducted two contrasts, one testing whether affirmations would reduce defensiveness, regardless of timing (pre-threat, post-threat: –1; control: +2), and one testing whether only the pre-threat affirmation would be effective (pre-threat: –2; control, post-threat: +1). As expected, the former was significant, $t(72) = 3.50$, $p = .001$, but not the latter, $t(72) = 1.43$, $p > .15$. In contrast to Study 1, participants in both the pre-threat and post-threat affirmation conditions responded less defensively to the negative test score than did participants in the control condition, $t(72) = 2.88$, $p = .01$, pre- versus control; $t(72) = 3.26$, $p = .002$, post- versus control. Among affirmed participants, the pre-threat affirmation was no more effective than the post-threat affirmation ($t < 1$).

Study 2b

In Study 2b, we sought to prompt some participants to initiate defensive responding just before they encountered the affirmation manipulation. According to our analysis, if participants initiate that defensive processing before being given a chance to self-affirm, the self-affirmation exercise will fail to

reduce the degree of defensive processing participants display. To do this, we replicated the control and post-threat affirmation conditions from Study 2a but replaced the pre-threat affirmation condition with a *foreshadowing* condition. Participants in this condition, like participants in the post-threat affirmation condition, were asked to complete a self-affirming essay immediately after scoring their own test. The only difference was that participants in the foreshadowing condition were told, just prior to writing their self-affirmation essay, what questions they would be answering after completing the "writing task." We presumed this would initiate defensive processing, which would fail to be undone by the affirmation exercise. By this reasoning, participants would show just as much defensiveness in the foreshadowing condition as participants in the no affirmation condition. The defensiveness of participants in the original post-threat condition, however, would be much lower. If, on the other hand, the reprocessing hypothesis is correct, then the foreshadowing manipulation should be irrelevant, and those in the foreshadowing and post-threat affirmations conditions should display less defensiveness than those in the control condition.

Method

Participants and design. In exchange for extra credit in their psychology and human development courses, 84 Cornell University undergraduates took part in the study. Participants were randomly assigned to a post-threat affirmation condition, a foreshadowing condition (in which the dependent measures were foreshadowed just prior to the post-threat affirmation), or a no affirmation control condition.

Procedure. The threat and measures of defensiveness were the same as those used in Study 2a, as was the control condition and the (nonforeshadowed) post-threat affirmation condition. A third group of participants was randomly assigned to a foreshadowing condition, which replaced the pre-threat affirmation condition. Participants in this foreshadowing condition, like participants in the post-threat affirmation condition, completed the self-affirmation manipulation after scoring their own test with the false answer key. How the foreshadowing condition differed was that the experimenter delivered the following lines just prior to the affirmation manipulation:

> You will have two tasks remaining today. First, you will complete a brief writing task. Second, we will ask you some follow-up questions to the test you completed today. We will ask you to make some test-specific judgments like how you think you would score on a comparably difficult alternate version of this test. Because the test is related to interpersonal perception ability, we'll ask you how many friends and acquaintances you have. And also, we'll have you rate yourself on a few domains, like intelligence.

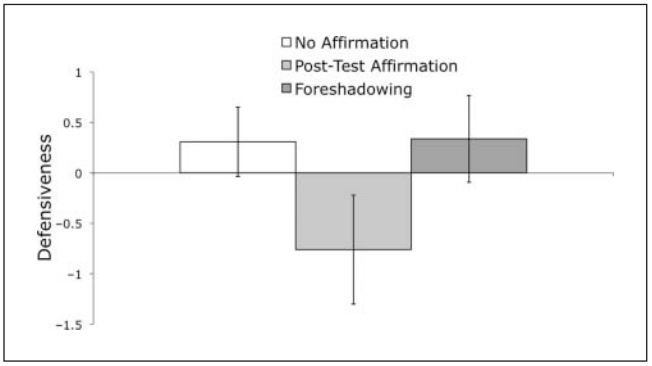

Figure 3. Defensiveness as a function of affirmation condition (Study 2b)
Values are z score composites and thus have no absolute meaning; 0 is not "no defensiveness."

It was expected that foreshadowing would put the wheels of defensive processing in motion and that, once these processes had been initiated, self-affirmations would be powerless to stop them. That is, participants in this condition would show just as much defensiveness as participants in the no affirmation condition and heightened levels relative to those in the post-threat condition.

Results

Perceptions of test performance. As in Study 2a, participants' ratings of their performance ($M = 3.8$) were significantly below the midpoint (5.0), suggesting that participants had experienced their performance as falling short of expectations, $t(83) = 7.35$, $p < .001$.

Measures of defensiveness. As in Study 2a, we standardized and averaged the responses to the measures of defensiveness. The means by condition are depicted in Figure 3. We again tested two contrasts: one was consistent with our timing hypothesis that the post-threat affirmation condition would be unique in reducing defensiveness (post-threat: –2; control, foreshadowing: +1). The other contrast examined the reprocessing-based prediction that the affirmation condition would equally reduce defensiveness compared to the control condition (post-threat, foreshadowing: –1; control: +2). The former contrast was significant, $t(79) = 2.12$, $p = .04$, but not the latter ($t < 1$). Participants were significantly more defensive in the foreshadowing condition than in the simple post-threat affirmation condition, $t(79) = 1.97$, $p = .05$. This comparison is the most direct test of whether foreshadowing eliminates the effect of post-threat affirmations. Nonaffirmed participants displayed marginally more defensiveness than those who were post-threat affirmed, $t(79) = 1.75$, $p = .08$, but no more defensiveness than those affirmed participants for whom the measures of defensiveness had been foreshadowed ($t < 1$).

Discussion

Studies 2a and 2b helped to reconcile our assertions about the timing of affirmation interventions with past work showing that affirmations can halt defensiveness even after participants are presented with a threat. In Study 2a, we showed, as in past work, that a post-threat affirmation can be successful in reducing defensiveness. However, in Study 2b, with the same threat and defensiveness measures, the very same self-affirmation was rendered ineffective—but only in the condition in which the pathways to defensiveness had been suggested to participants *prior* to the self-affirmation. Taken together, these results suggest that when participants form motivated, defensive conclusions prior to a self-affirmation, the affirmation will not undo these conclusions. Post-threat affirmations, therefore, may well be effective in blocking subsequent defensive processing but will not undo defensive conclusions when such processes have already taken place.

Study 3

The results of Studies 2a and 2b suggest that post-threat affirmations will be effective only if they are initiated before a threat response. If this is true, then we should be able to use the same logic to explain the apparent effectiveness of post-threat affirmations that have been documented in other domains. In Study 3, we returned to the origins of self-affirmation theory and examined its impact on cognitive dissonance after writing a counterattitudinal essay. Steele and Liu (1983) found that if participants self-affirmed *after* writing a counterattitudinal essay, they no longer defensively shifted their attitudes to justify having written the essay. According to our account, this post-threat affirmation was effective because attitude change had actually yet to occur, not because the self-affirmation undid an already drawn defensive reaction.

We had reasons to suspect that attitude change following counterattitudinal advocacy is not spontaneous but happens only upon presentation of an attitude questionnaire. Aronson, Blanton, and Cooper (1995) found that after writing an uncompassionate counterattitudinal essay, participants were willing to accept personality feedback that they were a compassionate person, but only if they had first been given a chance to shift their attitudes. If not first given the chance, they did not want to accept this feedback, which presumably would have highlighted the personal standard of compassion that participants had violated. Had attitude change been spontaneous, varying the timing of the presentation of the attitude measure would have had no effect.

If our hypothesis is correct, then foreshadowing our measures of defensiveness (the attitude measure) following the essay but prior to the self-affirmation essay (a *dissonance + foreshadowing* condition) the affirmation should be less effective than if no foreshadowing occurred (a *dissonance + no foreshadowing* condition). We added a third *no dissonance + foreshadowing* condition both as a no threat comparison group and to make certain that it was not simply the foreshadowing that prompted a defensive response. If the post-threat affirmation was effective in eliminating defensiveness but there was simply something about foreshadowing that led to more attitude change, then this should be true whether participants had first been made to experience dissonance or not. Thus, we predicted that those in the dissonance + foreshadowing condition would display more defensiveness than either those in the dissonance + no foreshadowing or the no dissonance + foreshadowing condition.

Method

Participants and design. In exchange for extra credit in their psychology and human development classes, 76 Cornell University undergraduates participated. All participants were asked to write a counterattitudinal essay and then to complete a self-affirmation exercise. Participants were randomly assigned to a dissonance + no foreshadowing, dissonance + foreshadowing, or no dissonance + foreshadowing condition. Of the participants, 8 refused to write the essay (4 in the dissonance + no foreshadowing condition, 3 in the dissonance + foreshadowing, and 1 in the no dissonance + foreshadowing), and 2 unfortunately astute participants identified the methodology as a "dissonance paradigm." These 10 participants were excluded from all analyses reported below, leaving 66 participants in our final set of analyses.

Procedure. The experimenter, intentionally reading off of a script, explained that the university's "Committee of Plans and Resources" was currently soliciting student feedback on whether to expend the resources to make all university buildings accessible to the physically disabled. It was explained that loopholes in state law had not required the university to bring its older buildings into compliance with current standards. It was said that at this point the committee was having participants write short, persuasive statements explaining why they "support or oppose a funding increase to help the physically disabled." Participants were told they would write a short essay that would be sealed in an envelope to be sent to the committee.

At this point, the experimenter looked up and put the paper from which she was reading aside to make it appear that she was going "off script." For those in the two dissonance conditions, the experimenter created the impression that writing the counterattitudinal essay was a matter of free choice by stating, "We actually already have enough essays written in favor of the funding increase. So we are asking participants if they wouldn't mind writing an essay against the funding increases. Is that OK?" For participants who agreed to write the essay, the experimenter gave them a form on which they were to write their essay and a letter-size Department of Psychology envelope. At the top of the

form was the university seal and the fictitious title of the committee, under which was written "On University Resolution 2007-0138. Position: OPPOSED." Just underneath, the form read,

> Thank you for your willingness to offer your opinion on this important issue affecting our university community. Please write a strong and convincing essay in the space below. The university should not increase spending for services for the physically disabled because . . .

This was followed by a blank page on which participants could write their essays. Each participant was told that upon completing the essay, he or she should fold and place it in the envelope and then seal and sign the envelope to indicate to the committee that the essay had not been tampered with.

In the no dissonance + foreshadowing condition, the ostensibly "off script" instructions provided to participants in the no dissonance + foreshadowing condition were intended to dispel any dissonance created by making it clear that the participant had no choice but to write the essay. Specifically, participants in this condition were told,

> Because we already have enough essays written in favor of the funding increase, we are just having participants write against the funding increase. Also, since we aren't giving you a choice, we are putting a star on your envelope so the committee knows that you were just assigned this position.

After writing the essay, participants in the two foreshadowing conditions were told,

> We have two tasks left for you to do. The first task is a writing task. Then, we will have you answer a few questions relevant to the first part of the experiment, such as to what extent you support or oppose the funding increase.

By describing the final questions as relevant only to "the first part of the experiment," the foreshadowing manipulation should not have communicated that there was a connection between the affirmation and the measures of defensiveness (Sherman et al., 2009). Participants in the no foreshadowing condition were not forewarned about the additional attitude-relevant questions. At this point, all participants completed the same self-affirmation manipulation used in Studies 2a and 2b. None of the values participants could affirm related to helping or showing compassion for others.

Finally, all participants answered to what extent they agreed with the statement, "The University should allocate more funds to improving facilities and services for the physically disabled" on a 17-point scale that ranged from 1 (*strongly disagree*) to 17 (*strongly agree*). The degree to which participants opposed the funding increase (and thus endorsed the counterattitudinal position) was taken as an indication of dissonance reduction or defensiveness.

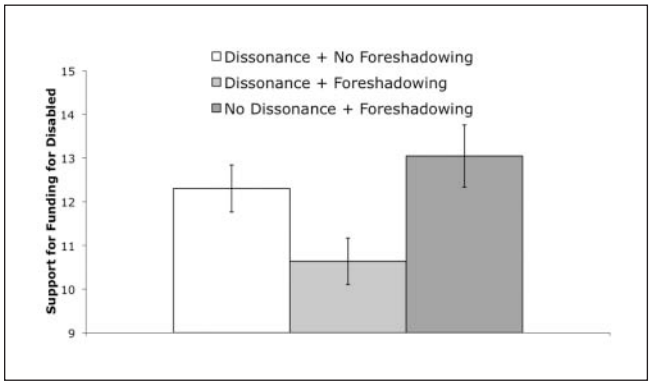

Figure 4. Support for the disabled as a function of dissonance foreshadowing condition (Study 3)
Lower values reflect greater defensiveness.

Results

We tested our timing hypothesis by contrasting the attitudes observed in the dissonance + foreshadowing condition (weighted –2) with the attitudes seen in the other two conditions (both dissonance + no foreshadowing and no dissonance + foreshadowing conditions weighted +1). The resulting statistical contrast was significant, $t(63) = 2.80$, $p = .01$ (see Figure 4). Conceptually replicating Study 2b, those in the dissonance + foreshadowing condition were less supportive of the physically disabled ($M = 10.6$) than those in the dissonance + no foreshadowing condition ($M = 12.3$), $t(63) = 2.00$, $p = .05$, and in the no dissonance + foreshadowing group ($M = 13.1$), $t(63) = 2.83$, $p = .01$. Attitudes in the latter two conditions did not differ ($t < 1$). We should note that these results are inconsistent with the view that the timing of affirmation does not matter. This view would predict that the three experimental groups would not differ, in that the affirmation intervention would have prevented attitude change regardless of foreshadowing.

Discussion

Study 3 used a very different type of defensiveness—attitude change following counterattitudinal advocacy—to test whether self-affirmations only at times appear to undo acts of defensiveness because the acts of defensiveness they blocked had actually yet to occur. A post-threat affirmation became ineffective once the to-be-measured method of defensiveness (attitude change) was foreshadowed for participants just prior to the self-affirmation.

The inclusion of the no dissonance + foreshadowing condition allowed us to rule out an alternative explanation that people in the foreshadowing condition simply had more time to consider their attitude toward the physically disabled, such that an initially positive reaction may have eventually given way to a more ambivalent stance. That is, perhaps with more thought, one would think of other worthy causes that would permit better or more efficient uses of the university's resources. This alternative predicts that foreshadowing would not merely prompt what looks like defensiveness after dissonance had been evoked but that it would have the same effect even when no dissonance had been evoked. The significant difference between the dissonance foreshadowing and the no dissonance foreshadowing conditions speaks against this possibility.[4]

General Discussion

We reason that if a self-affirmation satisfies a self-esteem need and this need had been heightened by a recent threat, then an affirmation should block subsequent acts of defensiveness. However, to the extent that motivated conclusions have already been reached, affirmations should not cause people to undo or "reprocess" such conclusions. This distinction (of timing in relation to a threat response rather than timing in relation to a threat experience) has to date been overlooked. Self-affirmations and defensiveness are interchangeable ways of alleviating threat (Tesser, 2000), and because the motivation to self-enhance is most active during threat (Tesser, 1988), an affirmation following defensiveness is unlikely to influence responding.

The results of the four studies herein confirm these predictions. We found that self-affirmations administered *prior* to the defensive response to threat attenuated subsequent defensive conclusions. When we used measures of defensiveness that assessed presumably direct and spontaneous reactions to a test (Study 1), a post-threat affirmation did not influence defensive responding. In circumstances in which post-threat affirmations blocked more indirect measures of defensiveness (Studies 2a and 2b) and attitude change in response to dissonance (Study 3), those affirmations became ineffective once participants were told before the affirmation about the possible defensive responses that would later be assessed. Without this foreshadowing condition, the timing of a threat response, and the effectiveness of an affirmation in relation to that response, would be unclear, and we suggest that this is one reason why the role of timing in affirmation effectiveness has not previously been identified. In other words, self-affirmation does not appear to occupy a privileged status among means of self-esteem restoration. After threat, people move to repair their self-integrity, and either an affirmation or a defensive response will do—and they will take whatever comes first. People do not necessarily reconsider their defensive responses in light of affirmation as past researchers have suggested (Cohen et al., 2000; Sherman et al., 2000).

The Foreshadowing Manipulation

The foreshadowing manipulation was meant to simply alert participants before the affirmation exercise of the defensiveness pathways that would ultimately be available to them and thus to start that process. But might the foreshadowing manipulation have instead led participants to take the affirmation task less seriously? If so, this might explain why the affirmation was no longer effective in the foreshadowing conditions.

To address this alternative, we returned to the actual affirmation essays that participants wrote in Studies 2b and 3. Two coders, blind to conditions and hypotheses, (a) rated the overall "affirming value" of each essay on a 1 to 7 scale based on a modified version of an affirmation coding scheme developed by Creswell et al. (2007) (rs = .83 and .93 for the two studies) and (b) counted the number of words in each essay (rs = 1.0). Contrary to an explanation that those in the foreshadowing condition took the affirmation task less seriously, they wrote essays that were just as affirming (ts = 1.13 and 1.48, ps > .14) and of no shorter length (ts < 1). Also, logistic regressions found that foreshadowing did not change the domain participants chose to affirm (ps > .20), and across all essays coders only twice observed any connection at all between the threatened domain and the content of the affirmation—once in each condition. In short, foreshadowing seemed to eliminate the tendency for affirmations to reduce subsequent defensiveness without changing the way people completed the affirmation task.

Why Wouldn't Timing Matter?

A consistent theme in social cognition is that timing of cognitive manipulations *does* matter (Von Hippel, Sekaquaptewa, & Vargas, 1995). A conceptual prime has the potential to affect perception of a stimulus only when experienced before, but not after, exposure to the target (Srull & Wyer, 1980). Schemas help to organize ambiguous information when learned of before exposure to the information but confer no benefit when learned of following exposure (Bransford & Johnson, 1973). People's expectations color their interpretation of bottom-up experience, but only when the expectation is put in place before the actual experience (Critcher & Dunning, 2009).

With all this research in mind, a careful reader may wonder how much news there is in our data that timing also matters in the realm of self-affirmation. Such a question is reasonable, but the extant literature does not necessarily guarantee that timing would still matter as one moves from the cognitive (e.g., expectations) to the motivational (e.g., self-affirmation). Not only does past evidence in self-affirmation, at first look, appear to suggest that timing would not matter (e.g., Cohen et al., 2000), there was reason to suspect that the reprocessing hypothesis (that timing would not matter) would hold water. For cognitive variables, such as expectations, to

influence perception, they must be in place before the stimulus is literally perceived because they shape the very encoding and interpretation of a stimulus. Self-affirmations do not directly color processing or encoding; instead, they exert a *subtractive* force, eliminating the influence of a motivation that would have pushed one toward desired perceptions. If one thinks of motivated conclusions as being bolstered or "propped up" by this motivation, then it would seem reasonable that subtracting out the force could diminish defensiveness, even after it is already instigated.

Where this intuition is in error, we suspect, is that even though motivation may be responsible for producing a perception, it is not the motivation that sustains that perception. As an illustration, threatened participants in Study 2b may have been motivated to call to mind scores of "friends" in defensively downplaying the implications of the threatening feedback. But once the motivation that spurred this intense search was removed, it does not follow that the acquaintances who were recalled would suddenly be forgotten. That is, taking away the motivation need not lead one to dismiss what past motivated reasoning has produced. More generally, subtracting the motivational influence that produces motivated distortions does not cause them to go away.

Theoretical and Applied Implications

Our data have at least four broad implications for self-affirmation theory and for the psychology of defensiveness more generally. First, because the moderating role of affirmations' timing seems to be pre-defensiveness or post-defensiveness instead of pre-threat or post-threat, post-threat affirmations may offer promise as a tool to determine whether defensiveness occurs spontaneously or merely once prompted by a dependent measure. For example, a comparison of Studies 1 and 2 suggests that people may be more likely to spontaneously engage in direct, as opposed to indirect, methods of alleviating threat. This is consistent with past research that suggests that those who have positive, affirming identity resources (an indirect means to self-repair) may have difficulty spontaneously relying on them unless the experimenter directs participants' attention to such self-esteem resources (Spencer, Josephs, & Steele, 1993). We suspect that one reason why research on the psychology of defense has uncovered such a variety of defensive strategies is not because people spontaneously use all of them but because they know *how* to use them once they are suggested by someone else.

Second, they suggest that defensive conclusions, once drawn, may often be crystallized rather quickly. In Studies 2b and 3, foreshadowed participants received the self-affirmation manipulation just seconds after being informed of what questions (the measures of defensiveness) they would answer after the writing task (the self-affirmation). The defensive conclusions that participants presumably drew in that short time span (e.g., "That test got me wrong; I have so many friends!") were not tempered by an immediately subsequent self-affirmation, suggesting a very short *critical period* for post-threat affirmations.

Third, and related, is to what extent a defensive response is cemented. Our data suggest that if a defensive response occurs spontaneously or is triggered before the motivation to be defensive is eliminated (whether through self-affirmation or perhaps the mere passage of time), the defensive conclusions reached through these processes will remain even once the motivation that spurred them is eliminated. Consistent with this conclusion, Dunning (2003) suggested that self-affirming defensive responses likely leave residuals of self-enhancement that over time combine to produce a better-than-average view of the self. Implicit in this argument is that motivated self-enhancement does not reverse itself even once self-integrity is restored.

Fourth, the crucial role of affirmations' timing may have important implications for the use of affirmations in applied or clinical settings. In general, practitioners are likely to have the most success with affirmation interventions when introduced just prior to a threat. But this situation may be complicated if one has already defensively downplayed the threat. Consistent with this notion, Epton and Harris (2008) note that self-affirmations may be more effective in encouraging new health-promoting behaviors than in discouraging health-deteriorating behaviors in which one already engages. If people are more likely to have defensively justified their bad habits than their failure to engage in a good habit, our research may explain this previously observed asymmetry. But we believe this picture is unnecessarily bleak, for an affirmation should be able to change a previously justified behavior when the affirmation is presented prior to a brand new appeal, one that uses a new persuasion tactic. For example, Armitage, Harris, Hepton, and Napper (2008) found that heavy cigarette smokers were more receptive to a novel anti-smoking message when they were first affirmed. When people affirm before exposure to this novel appeal, the affirmation can truly occur pre-defensiveness, giving it a higher chance of success.

Conclusion

In much of this article we have highlighted the perils of defensive responding (e.g., inaccurate self-knowledge, rejection of constructive feedback), but we do not wish to imply that defensive processes are always best to be suppressed. Rationalizing away a romantic interest's snub, downplaying the importance of one's own artistic ineptitude, and believing one will live forever may help avoid a life plagued by self-consciousness, low self-esteem, and high anxiety (Critcher et al., in press). At the same time, doing these things to excess may lead one to continue to pursue unattainable dates, unwisely invest one's inheritance promoting one's own bad artwork, and engage in high-risk behaviors. Determining the boons and banes of defensiveness processes

is a worthy task for continued research, though the present studies should aid practitioners in effectively harnessing the defensive-reducing power of self-affirmations.

Acknowledgments

We thank Jane Risen and Dennis Regan for their comments on a draft of the article. We thank Jill Fleischer, Katie McCrary, and Paige Weinger for their assistance with data collection. We thank Sally Apuzzo and Lisa Colton for coding the affirmation essays in Studies 2b and 3.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interests with respect to the authorship and/or publication of this article.

Financial Disclosure/Funding

The authors received the following financial support for the research and/or authorship of this article: This research was supported in part by a National Science Foundation Graduate Research Fellowship and a Yale University Mellon Research Grant awarded to Clayton Critcher and by National Institute of Mental Health Grant RO1 56072 and National Science Foundation Grant 0745806 awarded to David Dunning.

Notes

1. The answers are *cookie* and *box*.
2. A traditional control for this particular affirmation is to have participants write about why their least valued domain might be important to someone else. We worried that, to the extent that devalued domains might be domains on which one did not feel particularly competent, this "control" might serve as a threat and exaggerate any observed effects of the affirmation manipulation. To reduce this concern, we used a content-unrelated control that seemed unlikely to serve as either a threat or a self-affirmational resource for any participant.
3. Speaking to the comprehensiveness of the hypothesized contrasts, the residual variance is not significant in each study: $F(1, 96) = 3.01$, $p = .09$ (Study 1); $F < 1$ (Study 2a); $F < 1$ (Study 2b); $F < 1$ (Study 3). The marginally significant residual variance in Study 1 reflected the tendency for the post-threat affirmation to lead to marginally *more* defensiveness than when nonaffirmed and does not speak to the tenability of the competing reprocessing hypothesis.
4. Note that the inclusion of a no affirmation + dissonance condition would have allowed us to test whether the foreshadowing manipulation fully or only partially eliminated the impact of the affirmation, a more nuanced concern that was not central to the study's purpose, though note that this comparison is possible in Study 2b, which showed that foreshadowing eliminated the full impact of the affirmation.

References

Alicke, M. D., Klotz, M. L., Breitenbecher, D. L., Yurak, T. J., & Vredenburg, D. S. (1995). Personal contact, individuation, and the better-than-average effect. *Journal of Personality and Social Psychology, 68*, 804-825.

Armitage, C. J., Harris, P. R., Hepton, G., & Napper, L. (2008). Self-affirmation increases acceptance of health risk information among UK adult smokers with low socioeconomic status. *Psychology of Addictive Behaviors, 22*, 88-95.

Aronson, J., Blanton, H., & Cooper, J. (1995). From dissonance to disidentification: Selectivity on the self-affirmation process. *Journal of Personality and Social Psychology, 6*, 986-996.

Aspinwall, L. G. (1997). Future-oriented aspects of social comparisons: A framework for studying health-related comparison activity. In B. P. Bunk & F. X. Gibbons (Eds.), *Health, coping, and well-being: Perspectives from social comparison theory* (pp. 125-165). Mahwah, NJ: Erlbaum.

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin, 117*, 497-529.

Beauregard, K. S., & Dunning, D. (2001). Defining self-worth: Trait self-esteem moderates the use of self-serving trait definitions in social judgment. *Motivation and Emotion, 25*, 135-161.

Blanton, H., Cooper, J., Skurnik, I., & 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.

Bransford, J. D., & Johnson, M. K. (1973). Considerations of some problems of comprehension. In W. G. Chase (Ed.), *Visual information processing* (pp. 383-438). New York, NY: Academic Press.

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

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

Cohen, G., 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.

Costanzo, M., & Archer, D. (1989). Interpreting the expressive behavior of others: The Interpersonal Perception Task. *Journal of Nonverbal Behavior, 13*, 225-245.

Creswell, J. D., Lam, S., Stanton, A. S., Taylor, S. E., Bower, J. E., & Sherman, D. K. (2007). Does self-affirmation, cognitive processing, or discovery of meaning explain the cancer-related health benefits of expressive writing? *Personality and Social Psychology Bulletin, 33,* 238-250.

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

Critcher, C. R., Helzer, E., & Dunning, D. (in press). Self-enhancement via redefinition: Defining social concepts to ensure positive views of self. In M. Alicke & C. Sedikides (Eds.), *Handbook of self-enhancement.* New York: Guilford Press.

Dunning, D. (2003). The zealous self-affirmer: How and why the self lurks so pervasively behind social judgment. In S. J. Spencer, S. Fein, M. P. Zanna, & J. M. Olson (Eds.), *The Ontario symposium* (Vol. 9, pp. 45-72). Mahwah, NJ: Erlbaum.

Dunning, D. (2005). *Self-insight. Roadblocks and detours on the path to knowing thyself.* New York, NY: Psychology Press.

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

Epton, T., & Harris, P. R. (2008). Self-affirmation promotes health behavior change. *Health Psychology, 27*, 746-752.

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.

Hart, J., Shaver, P. R., & Goldenberg, J. L. (2005). Attachment, self-esteem, worldviews, and terror management: Evidence for a tripartite security system. *Journal of Personality and Social Psychology, 88*, 999-1013.

Klein, W. M., & Kunda, Z. (1989, March). *Motivated person perception: Justifying desired conclusions.* Paper presented at the meeting of the Eastern Psychological Association, Boston, MA.

Marks, G. (1984). Thinking one's abilities are unique and one's opinions common. *Personality and Social Psychology Bulletin, 10*, 203-208.

Markus, H. R., & Wurf, E. (1987). The dynamic self-concept: A social psychological perspective. *Annual Review of Psychology, 38*, 299-337.

McQueen, A., & Klein, W. M. P. (2006). Experimental manipulations of self-affirmation: A systematic review. *Self and Identity, 5*, 289-354.

Mednick, S. A. (1962). The associative basis of the creative process. *Psychological Review, 26*, 220-232.

Radcliffe, N. M., & Klein, W. M. P. (2002). Dispositional, unrealistic, and comparative optimism: Differential relations with the knowledge and processing of risk information and beliefs about personal risk. *Personality and Social Psychology Bulletin, 28*, 836-846.

Reed, M. B., & Aspinwall, L. G. (1998). Self-affirmation reduces biased processing of health-risk information. *Motivation and Emotion, 22*, 99-132.

Sedikides, C., Green, J. D., & Pinter, B. (2004). Self-protective memory. In D. R. Beike, J. M. Lampinen, & D. A. Behrend (Eds.), *The self and memory* (pp. 161-179). Philadelphia, PA: Psychology Press.

Sherman, D. K., & Cohen, G. L. (2006). The psychology of self-defense: Self-affirmation theory. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 38, pp. 183-242). San Diego, CA: Academic Press.

Sherman, D. K., Cohen, G. L., Nelson, L. D., Nussbaum, A. D., Bunyan, D. P., & Garcia, J. (2009). Affirmed yet unaware: Exploring the role of awareness in self-affirmation. *Journal of Personality and Social Psychology, 97*, 745-764.

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

Shira, I., & Martin, L. L. (2005). Stereotyping, self-affirmation, and the cerebral hemispheres. *Personality and Social Psychology Bulletin, 31*, 846-856.

Sivanthan, N., Molden, D. C., Galinsky, A. D., & Ku, G. (2008). The promise and peril of self-affirmation in de-escalation of commitment. *Organizational Behavior and Human Decision Processes*, *107*, 1-14.

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

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

Srull, T. K., & Wyer, R. S. (1980). Category accessibility and social perception: Some implications for the study of person memory and interpersonal judgments. *Journal of Personality and Social Psychology, 38*, 841-856.

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.

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

Taylor, S. E., & Lobel, M. (1989). Social comparison activity under threat: Downward evaluation and upward contacts. *Psychological Review, 96,* 569-575.

Tesser, A. (1988). Toward a self-evaluation maintenance model of social behavior. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 21, pp. 181-227). San Diego, CA: Academic Press.

Tesser, A. (2000). On the confluence of self-esteem maintenance mechanisms. *Personality and Social Psychology Review, 4*, 290-299.

Tesser, A., Martin, L. L., & Cornell, D. P. (1996). On the substitutability of self-protective mechanisms. In P. M. Golwitzer & J. A. Bargh (Eds.), *The psychology of action: Linking cognition and motivation to behavior* (pp. 48-68). New York, NY: Guilford.

Trope, Y., & Pomerantz, E. M. (1998). Resolving conflicts among self-evaluative motives: Positive experiences as a resource for overcoming defensiveness. *Motivation and Emotion, 22*, 53-72.

Von Hippel, W., Sekaquaptewa, D., & Vargas, P. (1995). On the role of encoding processes in stereotype maintenance. *Advances in Experimental Social Psychology, 27,* 177-254.

Wentura, D., & Greve, W. (2003). Who wants to be . . . erudite? Everyone! Evidence for automatic adaptation of trait definitions. *Social Cognition, 22*, 30-53.