

## Research Report

# Self-Affirmation Enhances Attentional Bias Toward Threatening Components of a Persuasive Message

William M.P. Klein<sup>1</sup> and Peter R. Harris<sup>2</sup>

<sup>1</sup>University of Pittsburgh and <sup>2</sup>University of Sheffield

**ABSTRACT**—We explored whether self-affirmation enhances attentional bias toward threatening elements of a persuasive message. Female alcohol consumers read an article linking alcohol to breast cancer and were then exposed supraliminally to threat and nonthreat words from the article (as well as threat and nonthreat words that did not appear in the article). Among moderately heavy drinkers who were not self-affirmed, there emerged an attentional bias away from the threatening words in the article—a result suggesting an avoidant response. However, among moderately heavy drinkers who were self-affirmed, there was a bias toward the threatening words. No attentional biases appeared for threat words not in the message, which suggested that the effect was threat specific. Moreover, no attentional biases were found among the heaviest drinkers. Self-affirmation may facilitate targeted implicit processing of threatening messages, although the effects could attenuate among individuals engaging in high levels of the behavior featured in the message.

People often respond defensively to messages that threaten their attributes, lifestyle, or prospects. For example, when reading messages that link their behavior to negative outcomes, they try to find fault with the messages (Kunda, 1987; Liberman & Chaiken, 1992). It is not surprising that people exhibit defensiveness, given that they appear to be motivated to hold positive self-views (Dunning, Heath, & Suls, 2004). This tendency is particularly problematic in that the individuals who are most at

risk are precisely those who engage in the most defensive responses (e.g., Croyle, Sun, & Hart, 1997).

When positive self-views can be maintained in the face of threat, people may be more evenhanded. Self-affirmation theory posits that people are motivated to sustain “self-integrity,” or perceptions that they are moral, consistent, and dependable (Steele, 1988). According to self-affirmation theory, messages designed to encourage behavior change threaten self-integrity and prompt the need to restore it. One way to reduce the threat is by denigrating the message. Self-affirmation theory proposes that self-integrity can be restored by affirming an equally important source of self-worth unrelated to the threat. The affirmation is not simply a distraction (see Correll, Spencer, & Zanna, 2004), but rather a means of offsetting the threat. A significant consequence is that self-affirmation in one domain of self-identity allows people to be more open to threats to self-integrity in another. The notion that self-affirmation promotes less defensiveness has been supported across a range of self-threats, including messages threatening lifestyle behaviors (e.g., Harris, Mayle, Mabbott, & Napper, 2007; Harris & Napper, 2005; Sherman, Nelson, & Steele, 2000).

How does self-affirmation influence engagement with a threatening message? This is less clear. Self-affirmation theory can be interpreted to suggest that self-affirmed individuals will devote more attentional resources than they otherwise would to the content of the message—contrary to non-self-affirmed individuals, who may reveal an attentional bias *away* from the threatening material. There is some initial evidence for this idea at a deliberate level: Participants in a study by Reed and Aspinwall (1998) were faster to navigate to risk-confirming than to risk-disconfirming information in a message. However, the existence of an implicit attentional bias has not been confirmed, as most studies simply measure attitude and belief change.

Address correspondence to William M.P. Klein, Behavioral Research Program, Division of Cancer Control and Population Sciences, National Cancer Institute, 6130 Executive Blvd., MSC 7326, Bethesda, MD 20892, e-mail: kleinwm@mail.nih.gov.

A second question is whether self-affirmation makes people more attentive to and engaged with the specific posed threat or increases overall vigilance. On the one hand, as argued earlier, self-affirmation is expected to enhance orientation toward and engagement with the material, especially its threatening elements. This should enhance responsiveness to material specifically in the message. On the other hand, self-affirmed individuals may show greater attentional bias to a range of threatening stimuli, much like anxious individuals (Frewen et al., 2008). The latter pattern seems unlikely, given that higher vigilance to irrelevant threatening information would be maladaptive (hence the association with anxiety), and would cast a cautionary note about the effects of self-affirmation. No research has addressed this question.

Whether self-affirmation is effective at all levels of risk also is unclear. Perhaps individuals at high risk have well-worn strategies for defensively resisting the threat, such that the effectiveness of self-affirmation would be reduced in these individuals (Sherman & Cohen, 2006; van Koningsbruggen & Das, 2008). Dissonance-arousing behaviors such as smoking and drinking may engage particularly stubborn defensive processes. Few studies recruit individuals at particularly high risk, so it is difficult to assess whether self-affirmation has a dose-response effect.

In the experiment we report here, female participants who had indicated that they consumed at least seven alcoholic beverages each week read an article that linked alcohol to breast cancer risk. Half were self-affirmed in advance. Next, all participants completed a visual-dot-probe task, an implicit measure of attentional bias to threat that is often used in the examination of anxiety (MacLeod, Mathews, & Tata, 1986). In such studies, anxious individuals who are exposed to threatening words or images display a greater attentional bias toward the words and images than nonanxious individuals (Bar-Haim, Lamy, Pergamin, Bakermans-Kranenburg, & van IJzendoorn, 2007; Frewen, Dozois, Joanisse, & Neufeld, 2008). We used the task in a different way by exposing participants to a threatening message (the article) and testing whether they showed an attentional bias to threatening words *in the message* (but no bias toward other threatening words). We predicted that non-self-affirmed participants would show an attentional bias away from threatening words in the message, which would reflect defensiveness. However, we expected self-affirmed participants to show an attentional bias *toward* these words. The sample comprised individuals whose level of alcohol consumption varied from moderately heavy to heavy, which made it possible to assess whether self-affirmation effects varied by risk level.

## METHOD

### Participants

Participants were 118 female undergraduates at the University of Pittsburgh who were fulfilling a research requirement and

participated individually. All reported, on a preexperimental questionnaire, that they consumed at least seven alcoholic beverages per week ( $M = 11.52$ ,  $SD = 4.51$ , range = 7–27).

### Procedure and Measures

Participants were greeted by the experimenter, signed informed-consent forms, and were then provided with a booklet. The first page constituted the self-affirmation manipulation and contained a list of values. Participants who had been randomly assigned to the self-affirmation condition ( $n = 54$ ) were instructed to select their most important value and to write a one-page essay about why it was important to them, including examples of situations in which it was expressed. Participants in the non-self-affirmation condition ( $n = 64$ ) were instructed to choose their least important value and write about why it might be important to someone else. This is a standard manipulation (McQueen & Klein, 2006). The top page of the booklet (which listed possible values) was always the same, so the experimenter was unaware of each participant's condition assignment.

Next, participants read an article (constructed by the authors) on a computer screen. This article linked alcohol consumption to breast cancer and recommended that women drink no more than one alcoholic beverage per day on average. Each person's reading time was recorded.

Participants then began the visual-dot-probe task, also on a computer. In each trial, a neutral word and a target word were presented together briefly and were very rapidly followed by the presentation of a dot. The dot appeared where either the target word or the neutral word had been. Participants responded by pressing the button corresponding to the dot's position. Faster reaction times to dots in the locations where threatening words had appeared would suggest allocation of attentional resources to threat.

Four sets of threat and neutral word pairings were presented in random order without replacement. Each set comprised 10 pairs, for a total of 40 pairs. The pairs included (a) threat and nonthreat words that had appeared in the article (threat-in items), (b) threat and nonthreat words that had not appeared in the article (threat-out items), (c) social anxiety and neutral words (social anxiety items), and (d) general-health-threat and neutral words (health items). The words were matched for length, number of syllables, and frequency in English, according to norms from Francis and Kučera (1982). The threat words were pilot-tested to be more negative than the neutral words.

Each word pair appeared twice, with the order of the words reversed the second time, for a total of 80 trials. Words were presented in black, 1-cm-high letters for 500 ms. On each trial, once the words disappeared, a 3-mm dot appeared in one position, where it remained until a button press. Participants indicated the dot's location by pressing the appropriate key (counterbalanced across conditions and handedness). Practice trials preceded the experimental trials. Participants then completed an unrelated questionnaire and were debriefed.

## RESULTS

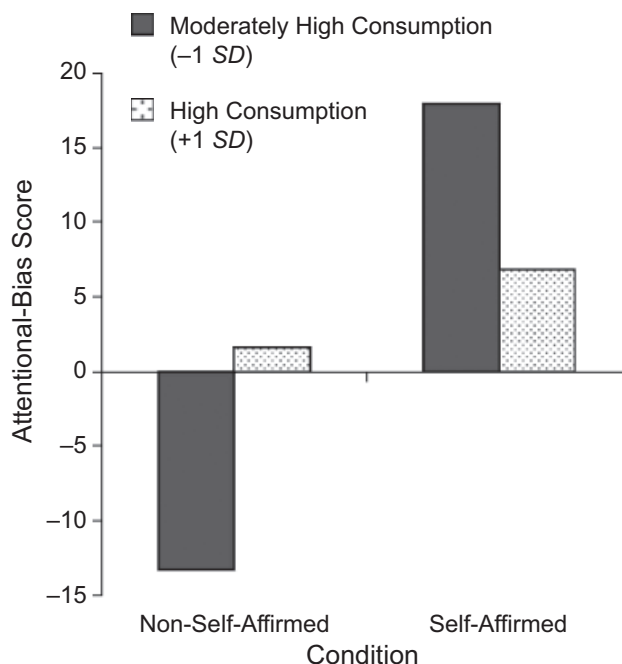
### Attentional Bias

If attention was directed toward threat words, response latency to the probe should have been faster when it appeared in the location of threat words. If attention was directed away from threat words, latency should have been slower when the probe appeared in the location of threat words. We created attentional-bias scores by subtracting response times for matched trials (i.e., the target word and the probe were in the same location) from response times for mismatched trials (i.e., the target word and the probe were in opposite locations). Thus, positive scores represented faster response times for matched trials, and negative scores represented slower response times for matched trials.

Attentional-bias scores were regressed on condition (self-affirmed or not self-affirmed, dummy-coded), amount of alcohol consumption, and their product. We conducted this analysis separately for threat-in, threat-out, health, and social anxiety items, as the study was not powered to detect a three-way interaction, given that many cells were not expected to differ. We computed a square-root transformation of the consumption variable (to reduce variability and skew) and mean-centered it before inclusion. Trials were excluded if the response was incorrect or if the response time was less than 150 ms or more than 2,000 ms. Only 2% of trials met one of these criteria.

As we had predicted, there was a significant effect of condition for threat-in words,  $\beta = .26$ ,  $t = 2.90$ ,  $p_{\text{rep}} = .98$ . Self-affirmed participants ( $M = 12.53$ ,  $SD = 40.51$ ) exhibited a stronger attentional bias toward threat-in words than did non-self-affirmed participants ( $M = -5.55$ ,  $SD = 27.84$ ). The bias score in the self-affirmation group was significantly greater than zero,  $t(53) = 2.27$ ,  $p_{\text{rep}} = .94$ ,  $r = .30$ ,<sup>1</sup> representing a true bias. Scores in the non-self-affirmed group were not significantly different than zero,  $t(63) = 1.60$ , n.s.

The regression analysis also revealed a main effect of consumption,  $\beta = .58$ ,  $t = 2.18$ ,  $p_{\text{rep}} = .94$ . Higher levels of consumption were associated with greater bias toward threat-in words. More important, the effect of condition was qualified by consumption,  $\beta = -.54$ ,  $t = 2.02$ ,  $p_{\text{rep}} = .92$ . As displayed in Figure 1, effects of self-affirmation were greater 1 standard deviation below the mean for alcohol consumption than at 1 standard deviation above. Participants who were moderately heavy drinkers showed an attentional bias *away* from the threatening words in the non-self-affirmed condition. Indeed, among participants whose level of consumption placed them below the mean-consumption level, attentional-bias scores were significantly *lower* than zero in the non-self-affirmed condition,



**Fig. 1.** Results of the regression analysis of attentional-bias scores for threatening words that had appeared in the article. The graph shows attentional-bias scores for non-self-affirmed and self-affirmed participants with reported alcohol consumption 1 standard deviation above and 1 standard deviation below the mean.

$M = -10.90$ ,  $t = -2.37$ ,  $p_{\text{rep}} = .95$ ,  $r = .35$ , and significantly *greater* than zero in the self-affirmed condition,  $M = 16.15$ ,  $t = 2.21$ ,  $p_{\text{rep}} = .93$ ,  $r = .35$ . No such differences emerged among the participants who had reported the heaviest alcohol consumption (see Fig. 1).

There were no significant effects in analyses for threatening words that did not appear in the article, health-related words, or social anxiety words,  $\beta s < .32$ ,  $t s < 1.13$ , n.s. Evidently, self-affirmation enhanced attention only to threatening words that had appeared in the article. Also, there were no group differences in response times for any word group,  $F s < 1.5$ , n.s.

### Reading Time

Heavy drinkers devoted more time to the article than did moderately heavy drinkers,  $\beta = .57$ ,  $t = 2.06$ ,  $p_{\text{rep}} = .92$ , but the analysis of reading times showed neither an effect of condition nor an interaction of condition and alcohol consumption,  $\beta s < .46$ ,  $t s < 1.65$ , n.s. Thus, the effects of self-affirmation concerned where attention was allocated rather than the overall extent of attention.

## DISCUSSION

Self-affirmed moderately heavy drinkers demonstrated an attentional bias toward threatening words in an article that linked alcohol consumption to breast cancer. In contrast, moderately heavy drinkers who were not self-affirmed displayed a bias away

<sup>1</sup>Effect sizes are reported as Pearson product-moment correlation coefficients ( $r$ ), following Rosenthal and Rosnow (1991). Cohen (1992) suggested that  $r$  values of .10, .30, and .50 represent small, medium, and large effects, respectively. Following Killeen (2005), we report  $p_{\text{rep}}$  (probability of replication) rather than  $p$  values.

from the threatening words. Thus, self-affirmation evidently altered the way in which this group engaged with the message. The self-affirmed individuals did not devote more time to the article, but rather processed it differently. The pattern was more apparent among moderately heavy alcohol consumers than among heavy alcohol consumers. These findings provide a clear demonstration—using an implicit measure of attentional allocation—of a basic change in orientation to threatening material induced by self-affirmation. It remains to be seen whether self-affirmation prevented a decrease or produced an increase in attentional bias, but either pattern would suggest a beneficial effect.

One important finding was that self-affirmed participants failed to show greater bias toward health- and threat-related words *not* in the article. Evidently, self-affirmation does not simply make people more vigilant, but rather prepares them to take opportunistic advantage of threatening material that is relevant to a specific threat. This is important because attentional bias to threat has been linked to anxiety (Bar-Haim et al., 2007; Frewen et al., 2008). Nonanxious individuals also show an attentional bias toward threat, just at a different threshold, which makes evolutionary sense given the survival value of processing threatening stimuli. Self-affirmation appears to be a contextual factor that can manipulate the threshold for bias—and in an adaptive way, given that attention appears to be oriented exclusively to relevant but not irrelevant threats.

One strength of this study is that it was conducted entirely with a high-risk sample that nevertheless possessed variability in level of risk. We observed that self-affirmation was effective for moderately heavy drinkers, but not for heavy drinkers. (Note, of course, that consumption was treated as a continuous variable in our analyses, so these descriptors are strictly heuristic.) Clearly there is a “sweet spot” when it comes to effects of self-affirmation on attention to threat, and the location of that spot may depend on many factors, such as the nature and magnitude of the threat. Stronger or different types of self-affirmations may be more effective among particularly high-risk individuals, among whom fear responses and defensive responses may cancel each other out (cf. Witte, 1992). Because only high-risk individuals participated, this study does not speak to how individuals respond to a threatening message when the threat is low in personal relevance.

Our findings begin to address the immediate cognitive effects of self-affirmation that likely precede other, more downstream consequences investigated in the past (e.g., risk perceptions and attitude change). To date, little is known about proximal cognitive effects of self-affirmation (see also van Koningsbruggen & Das, in press). We suspect that attentional bias occurs early in the chain of processes leading to nondefensive responding, which in turn leads to greater memory for and endorsement of the message (assuming the message is strong). When attentional bias is combined with greater perceptions of self-relevance and efficacy (Epton & Harris, 2008; Harris & Napper, 2005), the

ultimate effect should be beneficial changes in intentions and behavior.

Notably, the documented effects of self-affirmation on a variety of processes that lead to behavioral change suggest that it does not simply remove defensive motivation, but may exert effects at many stages. For example, self-affirmation may influence the association between risk perceptions and behavior (Klein & Monin, 2009). Future research also needs to determine the role of cognitive processes in self-affirmation examined here and elsewhere (Briñol et al., 2004; Reed & Aspinwall, 1998) relative to the role of affective processes influenced by self-affirmation, such as positive mood (e.g., Koole, Smeets, van Knippenberg, & Dijksterhuis, 1999), self-esteem (Sherman & Cohen, 2006), and other-directed feelings (Crocker, Niiya, & Mischkowski, 2008).

It is likely that self-affirmation exerts other influences on processing of threatening messages. These effects may include facilitation of comprehension and organization of material in the message, the framing of evenhanded hypotheses when evaluating the evidence, and deactivation of knowledge structures containing defensive beliefs. Moreover, self-affirmation is likely to have an array of motivational and emotional effects that, along with the cognitive effects enumerated here, produce the attitudinal and behavioral changes observed in the literature. We hope these findings will promote a better understanding of how self-affirmation works, as well as a more informed approach to how it can enhance the impact of consequential messages.

**Acknowledgments**—We thank Robin Bott, Alyssa Heintzelmann, Melinda Kelly, Samantha Maxwell, Lucy Napper, Yvonne Saadi, Michelle Scott, Jacob Shratter, and Olivia Stapsinski for assistance in collecting the data reported here, and Peter Furness for programming reaction time protocols.

## REFERENCES

- Bar-Haim, Y., Lamy, D., Pergamin, L., Bakermans-Kranenburg, M.J., & van IJzendoorn, M.H. (2007). Threat-related attentional bias in anxious and nonanxious individuals: A meta-analytic study. *Psychological Bulletin*, 133, 1–24.
- Briñol, P., Gallardo, I., Horcajo, J., de la Corte, L., Valle, C., & Díaz, D. (2004). Affirmation, confidence, and persuasion. *Psicothema*, 16, 27–31.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155–159.
- Correll, J., Spencer, S.J., & Zanna, M.P. (2004). An affirmed self and an open mind: Self-affirmation and sensitivity to argument strength. *Journal of Experimental Social Psychology*, 40, 350–356.
- 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.
- Croyle, R.T., Sun, Y., & Hart, M. (1997). Processing risk factor information: Defensive biases in health-related judgments and memory. In K.J. Petri & J.A. Weinman (Eds.), *Perceptions of*

- health and illness: Current research and applications (pp. 267–290). London: Harwood Academic.
- 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*, 74, 746–752.
- Francis, W.N., & Kučera, H. (1982). *Frequency analysis of English usage: Lexicon and grammar*. Boston: Houghton Mifflin.
- Frewen, P.A., Dozois, D.J.A., Joanisse, M.F., & Neufeld, R.W.J. (2008). Selective attention to threat versus reward: Meta-analysis and neural-network modeling of the dot-probe task. *Clinical Psychology Review*, 28, 307–337.
- Harris, P.R., Mayle, K., Mabbott, L., & Napper, L. (2007). Self-affirmation reduces smokers' defensiveness to graphic on-pack cigarette warning labels. *Health Psychology*, 26, 437–446.
- 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.
- Killeen, P.R. (2005). An alternative to null-hypothesis significance tests. *Psychological Science*, 16, 345–353.
- Klein, W.M.P., & Monin, M.M. (2009). When focusing on negative and positive attributes of the self elicits more inductive self-judgment. *Personality and Social Psychology Bulletin*, 35, 376–384.
- 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.
- Kunda, Z. (1987). Motivated inference: Self-serving generation and evaluation of causal theories. *Journal of Personality and Social Psychology*, 53, 636–647.
- Liberman, A., & Chaiken, S. (1992). Defensive processing of personally relevant health messages. *Personality and Social Psychology Bulletin*, 18, 669–679.
- MacLeod, C., Mathews, A., & Tata, P. (1986). Attentional bias in emotional disorders. *Journal of Abnormal Psychology*, 95, 15–20.
- McQueen, A., & Klein, W.M.P. (2006). Experimental manipulations of self-affirmation: A systematic review. *Self and Identity*, 5, 289–354.
- Reed, M.B., & Aspinwall, L.G. (1998). Self-affirmation reduces biased processing of health-risk information. *Motivation and Emotion*, 22, 99–132.
- Rosenthal, R., & Rosnow, R.L. (1991). *Essentials of behavioral research: Methods and data analysis* (2nd ed.). New York: McGraw-Hill.
- 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.
- 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.
- 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). New York: Academic Press.
- van Koningsbruggen, G.M., & Das, E. (2008, June). *When self-affirmation increases biased processing of threatening information: The moderating role of self-threat level*. Paper presented at the 15th general meeting of the European Association of Experimental Social Psychology, Opatija, Croatia.
- van Koningsbruggen, G.M., & Das, E. (in press). How self-affirmation reduces defensive processing of threatening health information: Evidence at the implicit level. *Health Psychology*.
- Witte, K. (1992). Putting the fear back into fear appeals: The extended parallel process model. *Communication Monographs*, 59, 329–349.

(RECEIVED 10/1/08; REVISION ACCEPTED 5/16/09)