Chapter 4

How Self-Affirmation Affects Implicit Health Information Processing*

People are often confronted with threatening information telling them they live an unhealthy lifestyle that poses a serious risk to their health. Although the wisest thing for people would be to respond by adopting healthy behaviors or stop risky behaviors, people may also refrain from accepting threatening health information, and consequently, persevere with their unhealthy behaviors (Sherman & Cohen, 2006). For instance, the more personally significant a health message is, the more people are likely to downplay the seriousness of the health risk, question the accuracy of the threatening information or evidence presented in the message, and process the information in a biased fashion (e.g., De Wit, Das, & Vet, 2008; Kunda, 1987; Liberman & Chaiken, 1992). Although such defensive processing of threatening health information is likely to keep worries at a safe distance, it may also prevent people from protecting their personal health. Self-affirmation theory (Steele, 1988) - a general theory about how people deal with self-threats - may help understand why people respond defensively to threatening health information.

According to self-affirmation theory (Steele, 1988), people are highly motivated to protect and maintain a global sense of self-integrity. Threatening information arouses this motive and people respond defensively to satisfy it (e.g., they downplay, minimize or avoid threatening health information). The theory further assumes that people are primarily concerned with their global sense of self-integrity. Consequently, they can also restore self-integrity by drawing upon alternative sources unrelated to the threat, such as reflecting upon an other important value (Steele, 1988). Thus, when threatened in one domain (e.g., health),

* This chapter is based on Van Koningsbruggen & Das (2008b) and Van Koningsbruggen, Das, & Roskos-Ewoldsen (in press).
people can restore their global self-integrity by affirming another important domain (e.g., intelligence). Such a “self-affirmation” restores global self-integrity, thereby reducing the need to respond defensively to the specific threat.

Although more and more studies show that theory-based manipulations of self-affirmation increase acceptance of threatening health information among people who normally tend to be defensive (e.g., Harris et al., 2007; Harris & Napper, 2005; Sherman et al., 2000), there is little direct empirical evidence regarding the cognitive processes that self-affirmation instigates (Sherman & Cohen, 2006). A better insight into these processes is pivotal in increasing our understanding of the processes that potentially mediate self-affirmation effects as well as their consequences for the design of health information. Most pressing is the issue of the origins and consequences of defensive processing at the implicit level: what makes people “turn off” from threatening health information, and is it true that self-affirmation makes people “turn on” to threatening health information again? Implicit processes are important to understand because they influence both how information is processed (Arpan, Rhodes, & Roskos-Ewoldsen, 2007) and subsequent behavior (Olson & Fazio, 2009). The present research provided a first test of the effect of self-affirmation on people’s implicit responsiveness to threatening health information by using an implicit, unobtrusive measure of information accessibility that provides more insight into the cognitive processes instigated by self-affirmation.

Self-Affirmation and Responses to Health Information

Previous self-affirmation research in the health domain has shown that self-affirmation can increase attendance to and acceptance of threatening health messages, increase perceptions of personal risk, strengthen intentions to take precautions, and promote behavior change regarding various health risks, such as smoking (Harris et al., 2007), excessive
caffeine consumption (Reed & Aspinwall, 1998; Sherman et al., 2000, Study 1), alcohol consumption (Harris & Napper, 2005), unsafe sex (Sherman et al., 2000, Study 2), insufficient fruit and vegetables consumption (Epton & Harris, 2008), and type 2 diabetes (Van Koningsbruggen & Das, in press). Some of these effects even remained stable over a period of one month (Harris & Napper, 2005).

Most studies focused on explicit persuasive outcomes (e.g., attitudes, intentions); less is known about the exact cognitive processes instigated by self-affirmation in response to threatening health information. To illustrate, Sherman and colleagues (2000) had coffee drinkers and non-coffee drinkers respond to information about a link between caffeine consumption and severe health problems. Typically, coffee drinkers respond defensively with the result that they accept such information to a lesser extent than non-coffee drinkers (e.g., Kunda, 1987). Self-affirmation reduced defensive responses among high-relevance participants: self-affirmed coffee drinkers were more likely to accept that caffeine consumption was linked to severe health problems and reported greater intentions to reduce their caffeine consumption than coffee drinkers who had not self-affirmed. Using a similar paradigm, self-affirmation made high-relevance participants read risk-confirming information earlier than risk-neutral or risk-disconfirming information (Reed & Aspinwall, 1998). However, contrasting previous findings, self-affirmed high-relevance participants reported lower intentions to reduce their caffeine consumption than their non-affirmed counterparts.

One recent study reported more direct evidence regarding the cognitive processes instigated by self-affirmation in response to threatening health information, by drawing on dual process models of persuasion (e.g., Petty & Cacioppo, 1986). When self-affirmation reduces defensive processing, people should differentiate between weak and strong arguments in a message and recognize the merits and demerits of these arguments (Petty & Cacioppo, 1986). Consistent with this reasoning, self-affirmation made people more sensitive to the
quality of the arguments in a health message (Van Koningsbruggen & Das, 2008b). These findings suggest that self-affirmation gives people the resources to deal with threatening health information, thus enabling them to process this information in a more open-minded fashion.

Most pressing is the absence of research on self-affirmation in the area of less controlled, automatic cognitive responses to threatening health information. Since many self-regulatory processes occur automatically and outside awareness (Bargh & Chartrand, 1999; Sherman & Cohen, 2006), it is important to test whether the proposed resource function of self-affirmation may also extend to this area, and if it increases responsiveness to threatening health information at the implicit level. Theoretically, self-affirmation is presumed to diminish self-integrity concerns and stimulate integration of the threat into the self-system (Sherman & Cohen, 2006). Thus, self-affirmation should make it easier to encode threatening aspects of health information because these aspects no longer pose a threat to people’s self-integrity. Because information that is better encoded should be easier retrieved from memory (i.e., more accessible; Ashcraft, 2006; Carlston & Smith, 1996; Higgins, 1996), we measured the accessibility of threat-related cognitions with an unobtrusive lexical decision task in the present research. In this task, participants have to decide as quickly and accurately as possible whether a word that appears on the screen is an existing word or a non-word. Among the existing words are both words related to the threat and neutral words. Greater accessibility of threat-related cognitions is indicated by faster response latencies to threat-related words. When self-affirmation promotes responsiveness to health information at the implicit level, it should increase the accessibility of threat-related cognitions. As far as we know, the present research is the first to test the impact of self-affirmation on the accessibility of threat-related health cognitions.
Although self-affirmation may promote increases in cognitive openness to threatening information, recent research that examined the impact of self-affirmation on health information processing at the explicit level suggests that this beneficial effect of self-affirmation may be limited to conditions in which people experience a moderate threat to the self (Van Koningsbruggen & Das, 2008b). Under low and high self-threat conditions, self-affirmation decreased extensive, careful processing of threatening health information, as indicated by a decrease in participants’ sensitivity to the quality of the arguments in the health information. Only under moderate self-threat conditions, self-affirmation increased extensive, careful health information processing. These effects on the explicit level of information processing may well originate from people’s implicit responses to threatening health information. Thus, in line with these previous findings, we propose that the effect of self-affirmation on people’s responsiveness to threatening health information at the implicit level is likely to be moderated by self-threat level: whereas self-affirmation may promote responsiveness to health information at the implicit level under moderate self-threat conditions, self-affirmation may impede implicit responsiveness under low and high self-threat conditions.

**Overview of the Present Studies**

In this research we examined the effect of self-affirmation on people’s implicit responsiveness to threatening health information under varying levels of threat to the self (low, moderate, and high) with a lexical decision task that measured the accessibility of threat-related health cognitions. We predicted that self-affirmation would increase accessibility of threat-related cognitions under moderate self-threat conditions, and decrease accessibility of threat-related cognitions under low and high self-threat conditions. Self-threat level was conceptualized as participants’ vulnerability to a health risk.
Previous research has shown that different self-threat levels can be created by employing different methods to classify participants’ vulnerability to a health risk (Van Koningsbruggen & Das, 2008b). Classifying vulnerability on participants’ actual behavior (such as distinguishing between coffee drinkers and non-coffee drinkers) allows for testing the impact of self-affirmation under low self-threat (i.e., non-coffee drinkers or people not at-risk) and moderate self-threat conditions (i.e., coffee drinkers or people at-risk). In Study 4.1 we employed this approach to test the impact of self-affirmation on people’s implicit responsiveness to threatening health information under low and moderate self-threat conditions. In Study 4.2, we employed a manipulation of vulnerability by providing participants with false feedback regarding their vulnerability level to a health risk. Recent research has shown that a manipulation of vulnerability allows for testing the impact of self-affirmation under moderate and high self-threat conditions (Van Koningsbruggen & Das, 2008b).

**Study 4.1**

In this study, we employed the “caffeine consumption paradigm” that has been frequently used in previous research on defensive processing of health information (e.g., see Block & Williams, 2002; Kunda, 1987; Liberman & Chaiken, 1992; Reed & Aspinwall, 1998; Sherman et al., 2000). Moreover, because previous studies that used this paradigm provided contradictory evidence regarding the impact of self-affirmation on intentions (Reed & Aspinwall, 1998; Sherman et al., 2000), we explicitly assessed intentions to take precautions and perceptions of health message quality. In doing so, we meet the call for delineating how cognitive processes influence health self-regulatory efforts (Williams, Wasserman, & Lotto, 2003).
How Self-Affirmation Affects Implicit Health Information Processing

Following previous research (e.g., Sherman et al., 2000; Van Koningsbruggen & Das, 2008b), we recruited coffee drinkers (i.e., moderate self-threat condition) and non-coffee drinkers (i.e., low self-threat condition) and presented them with health information linking caffeine consumption to severe health problems. Prior to reading this information, we manipulated self-affirmation by allowing participants to affirm a value that was either personally important (self-affirmation) or unimportant (no affirmation) to them. All participants then completed an unobtrusive lexical decision task to assess the accessibility of threat-related cognitions. Subsequently, they reported their perceptions of message quality and their intentions to take precautions. We predicted that among participants in the moderate self-threat condition (i.e., coffee drinkers), relative to participants in the low self-threat condition (i.e., non-coffee drinkers), self-affirmation would increase accessibility of threat-related cognitions, perceptions of message quality, and intentions to reduce caffeine consumption.

Method

Design and Participants

The hypotheses were tested in a 2 (self-threat level: low vs. moderate) x 2 (self-affirmation status: non-affirmed vs. self-affirmed) between-participants factorial design. A total of 84 university students participated in the experiment, of which 20 males and 64 females with a mean age of 23.62 years ($SD = 3.66$ years). As compensation for their participation, participants could take part in a lottery in which they could win gift vouchers (€25, approximately $35).

Procedure and Materials

Participants were informed that they would participate in several separate computerized studies: one about consumption patterns concerning several beverages, one
about values, one assessing student opinions of scientific articles, and the other about word recognition.

Self-threat level. After the introduction, participants were presented with several questions designed to identify their coffee drinking behavior (e.g., coffee drinker yes/no, if yes: whether they consumed regular or decaffeinated coffee etc.). To bolster our cover story, similar questions were asked about several other beverages. The sample consisted of 47 coffee drinkers (i.e., moderate self-threat condition) and 37 non-coffee drinkers (i.e., low self-threat condition). Participants then continued with the “value study” that contained the manipulation of self-affirmation.

Self-affirmation. The manipulation of self-affirmation was based on a well-established procedure (e.g., Harris & Napper, 2005; Sherman et al., 2000), whereby participants complete a brief writing assignment. Participants were randomly assigned to either the self-affirmed status condition (n = 40) or the non-affirmed status condition (n = 44). They were first presented with the six values of the Allport-Vernon-Lindzey study of Values (Allport et al., 1960). The values listed were: science, business, art, social, politics, and religion. In the self-affirmed status condition, participants were asked to choose their most important value and to write about why it was important to them and to describe a specific occasion when it had been particularly important. In the non-affirmed status condition, they were asked to choose their least important value and to write about why the value might be important to the average student.

Health message. After the manipulation of self-affirmation, participants read a fictitious article, entitled “The effects of caffeine on health”, which was supposedly published in “The Journal of Medicine” (based on Block & Williams, 2002). The first part of the article presented several examples of products containing caffeine. It was stressed that the greatest part of people’s caffeine intake could be related to their coffee consumption. The second part
described possible health problems related to caffeine consumption (e.g., insomnia, restlessness, high blood pressure, cardiovascular disease). Moreover, it was stated that research suggested that people who consume high doses of caffeine, for example, coffee drinkers, are more likely to suffer a heart attack or a stroke.

After participants read the health message, they were asked to participate in the study on word recognition (i.e., the lexical decision task). Participants then completed the perceived message quality and intention measures. After completing these measures, participants were probed for suspicion about the purpose of the study, and then were extensively debriefed. None of the participants guessed any aspect of the true purpose of the study, and none of them reported a suspicion that the studies were related.

**Dependent Variables**

*Accessibility of threat-related cognitions.* Participants were informed that a sequence of words would appear individually in the center of the screen. They were asked to decide as quickly and accurately as possible whether the presented word was an existing Dutch word or a non-word by pushing either the “A” (for existing words) or the “L” (for non-words) on the keyboard. Each trial first presented a fixation point in the center of the screen that was replaced by the stimulus after 1 second. The next trial was initiated 500 ms after participants pressed one of the keys. In total, participants responded to 40 words, consisting of 20 existing Dutch words and 20 non-words. In reality, we were not interested in reaction times (RT) to non-words, but only in the difference in RT to existing target words between conditions. Among the existing words, 5 were threat-related target words from the health message participants had just read, pertaining to the adverse health problems of caffeine consumption (blood pressure, heart attack, heart disease, restless, sick; all single words in Dutch), the remaining items reflected control words (e.g., apartment, mechanic, motorway). The threat-related words were selected on the basis of a pilot study in which participants ($N = 41$, not
taking part in the actual experiment) rated on a 7-point scale (1 = not at all, 7 = very much) to what extent these and 10 other words pertained directly to health problems related to caffeine consumption. The five words selected were the ones rated highest on this measure ($M = 5.15$, $SD = 1.99$). All words were matched on word length across word type categories, randomly presented, and preceded by four practice trials.¹

*Perceived message quality.* To assess participants’ perceptions of message quality, they responded to the question “The evidence linking caffeine consumption and health problems is reliable” (1 = strongly disagree, 7 = strongly agree). Higher scores indicate higher perceptions of message quality.

*Intentions to reduce caffeine consumption.* Participants rated two questions that examined their intentions to reduce their caffeine consumption on a 7-point scale (based on Block & Williams, 2002; “I intend to cut down my caffeine consumption” and “I am convinced that I will reduce the amount of caffeine I consume”; 1 = strongly disagree, 7 = strongly agree; $r = .79$, $p < .001$). Higher scores indicate greater intentions to reduce caffeine consumption.

**Results**

*Accessibility of Threat-Related Cognitions*

The computer recorded both the response (i.e., word or non-word) and response time (in milliseconds) per presented item for each participant. Incorrect responses were excluded from the analysis (2.86% across the experimental trials, which were evenly distributed across conditions). Following the recommendations of Ratcliff (1993), harmonic means of RT for threat-related words and control words were subjected to a 2 (self-threat level: low vs. moderate) x 2 (self-affirmation status: non-affirmed vs. self-affirmed) x 2 (word type: threat-related vs. control) mixed design analysis of variance (ANOVA) with repeated measures on
the third factor. The analysis revealed a main effect of word type, \(F(1, 80) = 17.49, p < .001, \eta_p^2 = .18\). Participants responded faster to threat-related words \((M = 775 \text{ ms})\) than to control words \((M = 837 \text{ ms})\). In addition, the analysis revealed the predicted interaction effect between self-threat level, self-affirmation, and word type, \(F(1, 80) = 7.79, p = .007, \eta_p^2 = .09\). Simple effects analysis indicated that in the moderate self-threat condition, self-affirmed participants responded faster to threat-related words than non-affirmed participants, \(F(1, 80) = 4.69, p = .033, \eta_p^2 = .06\). No effect of self-affirmation was found in the low self-threat condition, \(F < 1\). Also, no self-affirmation effects occurred on the responses to the control words (moderate self-threat condition, \(F < 1\); low self-threat condition, \(F[1, 80] = 1.19, ns\)). Relevant means are displayed in Table 4.1.

Table 4.1. Mean harmonic mean RT in milliseconds as a function of self-threat level, self-affirmation status condition, and word type (Study 4.1)

<table>
<thead>
<tr>
<th>Condition</th>
<th>Low self-threat (non-coffee drinkers)</th>
<th>Moderate self-threat (coffee drinkers)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Non-affirmed</td>
<td>Self-affirmed</td>
</tr>
<tr>
<td>Threat-related</td>
<td>753&lt;sub&gt;a&lt;/sub&gt;</td>
<td>761&lt;sub&gt;a&lt;/sub&gt;</td>
</tr>
<tr>
<td>Control</td>
<td>849&lt;sub&gt;a&lt;/sub&gt;</td>
<td>786&lt;sub&gt;a&lt;/sub&gt;</td>
</tr>
</tbody>
</table>

*Note.* Means with different subscripts (in rows, for each self-threat condition) differ significantly from each other \((p < .05)\).

**Perceived Message Quality**

A 2 (self-threat level) x 2 (self-affirmation status) ANOVA on perceived message quality revealed a significant interaction between self-threat level and self-affirmation, \(F(1, 80) = 7.04, p = .01, \eta_p^2 = .08\). Simple effects analysis revealed that in the moderate self-threat condition, self-affirmed participants reported higher perceptions of message quality \((M = \ldots\).
4.56, SD = 0.64) than non-affirmed participants (M = 3.90, SD = 0.91), F(1, 80) = 5.60, p = .02, ηp² = .07. No effect of self-affirmation was found in the low self-threat condition (Mself-affirmed = 3.69, SDself-affirmed = 0.95; Mnon-affirmed = 4.17, SDnon-affirmed = 1.20; F[1, 80] = 2.15, ns).

**Intentions to Reduce Caffeine Consumption**

A 2 (self-threat level) x 2 (self-affirmation status) ANOVA on intentions to reduce caffeine consumption revealed a significant interaction between self-threat level and self-affirmation, F(1, 80) = 6.03, p = .016, ηp² = .07. Simple effects analysis revealed that in the moderate self-threat condition, self-affirmed participants reported greater intentions to reduce their caffeine consumption (M = 3.02, SD = 1.73) than non-affirmed participants (M = 2.10, SD = 0.97), F(1, 80) = 5.42, p = .022, ηp² = .06. No effect of self-affirmation was found in the low self-threat condition (Mself-affirmed = 1.85, SDself-affirmed = 0.55; Mnon-affirmed = 2.42, SDnon-affirmed = 1.38; F[1, 80] = 1.54, ns).

**Relationship between Implicit and Explicit Measures**

Following Baron and Kenny (1986), we tested whether the difference in RT to threat-related words mediated the impact of self-affirmation on perceived message quality or intentions among participants in the moderate self-threat condition (controlling for overall speed of responding; cf. Rhodes, Roskos-Ewoldsen, Edison, & Bradford, 2008). Self-affirmation predicted RT to threat-related words, B = -48.09, t = -2.66, p = .011, and also perceived message quality, B = 0.34, t = 2.97, p = .005, and intentions, B = 0.47, t = 2.14, p = .038. However, RT to threat-related words did not significantly predict perceived message quality (p = .38) or intentions (p = .19). Thus, no further analyses were conducted since not all conditions for mediation were met.²

Finally, we performed partial correlation analyses between RT to threat-related words and the perceived message quality and intention measures among participants in the moderate self-threat condition. As can be seen in Table 4.2, RT to threat-related words were negatively
related to perceived message quality among self-affirmed participants. Thus, participants with faster response latencies to threat-related words reported higher perceptions of quality of the message linking caffeine consumption and health problems. This finding, in conjunction with the results on the lexical decision task, suggests that increased accessibility of threat-related cognitions is related to less defensive responding to threatening health information.

### Table 4.2. Partial correlations between RT to threat-related words and perceived message quality and intentions for participants in the moderate self-threat condition (Study 4.1)

<table>
<thead>
<tr>
<th>Condition</th>
<th>Variable</th>
<th>$r$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RT Threat-related words$^a$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-affirmed status</td>
<td>Perceived message quality</td>
<td>.15</td>
<td>.57</td>
</tr>
<tr>
<td></td>
<td>Intentions</td>
<td>-.13</td>
<td>.60</td>
</tr>
<tr>
<td>Self-affirmed status</td>
<td>Perceived message quality</td>
<td>-.40</td>
<td>.046</td>
</tr>
<tr>
<td></td>
<td>Intentions</td>
<td>-.01</td>
<td>.97</td>
</tr>
</tbody>
</table>

$^a$ Controlling for overall speed of responding.

### Discussion

Study 4.1 provided a first test of the impact of self-affirmation on people’s implicit responsiveness to threatening health information. The results revealed a highly consistent pattern of findings across implicit and explicit measures. As predicted, it was found that self-affirmation increased participants’ accessibility of threat-related cognitions, positively affected perceptions of message quality and increased intentions to take precautions under moderate self-threat conditions.

While there is some evidence that self-affirmation increases extensive, careful processing of threatening health information on a conscious, explicit level (Van Koningsbruggen & Das, 2008b), the present study suggests that when self-affirmation...
diminishes concerns about self-integrity, threatening health information becomes easier to retrieve from memory. This also fits in nicely with and supplements previous research showing that self-affirmation makes people read risk-confirming information earlier (Reed & Aspinwall, 1998). Because information that is more accessible will be more likely encoded and recalled (Ashcraft, 2006; Carlston & Smith, 1996; Higgins, 1996), the present results are consistent with the idea that self-affirmation stimulates integration of the threat into the self-system (Sherman & Cohen, 2006). Consequently, because the threatening aspects of the health information no longer pose a threat to people’s self-integrity it should reduce the need, for example, to derogate the health information. Consistent with this, we found that the more accessible threat-related cognitions were for self-affirmed participants in the moderate self-threat condition, the more enhanced their perceptions of quality of the message linking caffeine consumption and health problems were.

Additionally, whereas previous studies that employed the caffeine consumption paradigm provided contradictory results regarding the impact of self-affirmation on intentions (Reed & Aspinwall, 1998; Sherman et al., 2000), the present findings add to the growing evidence that self-affirmation can promote intentions to take precautions. Although we did not assess whether participants actually translated their intentions into actions, recent findings suggest that self-affirmation can promote behavior change. One study that examined screening behavior demonstrated that self-affirmation increased risk test taking among people for whom the health information was highly relevant (Van Koningsbruggen & Das, in press), and another study showed that self-affirmation increased fruit and vegetables consumption (Epton & Harris, 2008). However, future studies are needed to extend these findings to other behaviors. Additionally, while the accessibility of the threatening information was not related to behavioral intentions, this is consistent with a broad array of research demonstrating that construct and attitude accessibility play a greater role in the prediction of spontaneous
behaviors rather than deliberative behavior or behavioral intentions (Olson & Fazio, 2009). Testing the impact of self-affirmation on spontaneous health behaviors constitutes an important agenda for future studies.

Whereas previous research showed that self-affirmation decreases extensive, careful information processing when people are not particularly threatened (Briñol et al., 2007; Van Koningsbruggen & Das, 2008b, Study 2), in the present study we found no self-affirmation effects under low self-threat conditions. This may suggest that self-affirmation differently affects explicit and implicit measures of information processing when facing a low threat to the self. Alternatively, this may point to the fact that a manipulation of argument quality is necessary in order to detect a decrease in information processing under low self-threat conditions. Consistent with this latter argument, studies that did not manipulate argument quality similarly did not obtain self-affirmation effects on information processing under low self-threat conditions (e.g., Van Koningsbruggen & Das, 2008b, Study 1). Future research should address this issue more specifically.

Although the present findings do show that self-affirmation increases implicit responsiveness to threatening health information among people for whom the information is highly relevant, we predicted that this beneficial effect of self-affirmation would be limited to conditions in which people experience a moderate self-threat. When severely threatened, recent research demonstrated that self-affirmation backfires on the (explicit) processing of threatening health information (Van Koningsbruggen & Das, 2008b). Study 4.2 examined whether self-affirmation would similarly backfire on implicit responsiveness to threatening health information following a severe threat to the self.
Study 4.2

In this study we tested the effect of self-affirmation on people’s implicit responsiveness to threatening health information under moderate and high self-threat conditions. Following previous research (Van Koningsbruggen & Das, 2008b), we manipulated vulnerability to create moderate and high self-threat conditions by providing participants with false feedback regarding their vulnerability level to a health risk. To ensure that the findings of Study 4.1 were not health-topic specific, we used a different health topic: health risks associated with toxic ingredients in personal care products. Study 4.1 also included a different self-affirmation manipulation in which participants had to complete a value scale that allowed them to affirm a value that was either personally important (self-affirmation) or unimportant (no affirmation) to them. It was hypothesized that self-affirmation would increase the accessibility of threat-related cognitions under moderate self-threat conditions, and would decrease the accessibility of threat-related cognitions under high self-threat conditions. Finally, we assessed implicit affect in Study 4.2 to test whether the effects of self-affirmation may be explained by changes in affect. Although self-affirmation typically has no effect on explicit measures of affect (see Sherman & Cohen, 2006; Steele, 1988), there is one isolated example in which self-affirmation has been shown to influence implicit affect (Koole et al., 1999). Ruling affect out as an alternative explanation for the observed effects is important as it could plausibly affect information processing in similar ways (e.g., Raghunathan & Trope, 2002).

Method

Design and Participants

The hypotheses were tested in a 2 (self-threat level: moderate vs. high) x 2 (self-affirmation status: non-affirmed vs. self-affirmed) between-participants factorial design.
Participants were 79 university students (31 males, 48 females; \(M_{\text{age}} = 22.11\) years, \(SD_{\text{age}} = 4.85\) years). In exchange, they received €3.50. Participants were randomly assigned to experimental conditions.

Procedure and Materials

Participants were informed that they would participate in several separate computerized studies: one about the use of personal care products, one about values, one assessing student opinions of scientific articles, and the other about word recognition. The first part of the experiment was aimed at identifying participants’ most and least important values to manipulate self-affirmation. Participants were presented with the six values of the Allport-Vernon-Lindzey study of Values (AVL; Allport et al., 1960) which they had to rank according to their personal importance. The values listed were: science, business, art, social, politics, and religion. After participants ranked the values, they engaged in the manipulation of self-threat level.

Self-threat level. To manipulate self-threat level, participants were presented with several questions to identify the personal care products they used (i.e., what brands of deodorant, shampoo, and shower gel they used). After this, they were told that some brands contain toxic ingredients that can cause serious health problems. Participants were told about a test that analyzed whether the ingredients of the brands they used could cause these health problems. They were led to believe that the computer analyzed the details they provided earlier and were instructed to go to the next screen to see the result of this analysis (cf. Van Koningsbruggen & Das, 2008b). In the moderate self-threat condition, they saw a square, filled with a bright green color that displayed: “Your risk of developing health problems is quite low due to the brands you use.” In the high self-threat condition, participants saw a square, filled with a bright red color that displayed: “Your risk of developing health problems
is quite high due to the brands you use.” Participants then proceeded with the self-affirmation manipulation.

Self-affirmation. The manipulation of self-affirmation was based on a well-established procedure (cf. Koole et al., 1999; Sherman et al., 2000; Tesser & Cornell, 1991), whereby participants receive a 10-item AVL-subscale (Allport et al., 1960) either matching their previously indicated most important value (self-affirmed status condition) or least important value (non-affirmed status condition). Each of the ten items consisted of a statement with two possible answers. One of the answers reflected participant’s most important value (self-affirmed status condition) or least important value (non-affirmed status condition). The other answer was filler.

Health message. Participants read a message about toxic ingredients in personal care products. The message stated that many products contain toxic, hormone-disturbing ingredients that damage people’s health. It was described that the increased use of personal care products probably caused an increase in chronic diseases and allergies. Furthermore, participants were told that - when toxic personal care products were used regularly - the negative effects on people’s health were assumed to be “persistent” and “accumulative”, and possibly increased the risk of cancer.

After reading the health message, participants completed the dependent measures. They were then probed for suspicion about the purpose of the study, and then were extensively debriefed. None of the participants guessed any aspect of the true purpose of the study, and none of them reported a suspicion about the vulnerability feedback or that the studies were related.

Dependent Variables

Manipulation checks. To confirm the success of the self-threat level manipulation, participants rated four items that measured their perceived vulnerability to health problems
related to the use of personal care products on a 7-point scale (e.g., “The probability that I will experience health problems due to the personal care products I use is high”; 1 = strongly disagree, 7 = strongly agree; α = .83). Higher scores indicate higher perceptions of vulnerability (i.e., self-threat).

The validity of the self-affirmation manipulation was assessed by counting the number of times participants endorsed the manipulated value. A score of 1 was given when participants chose the manipulated value (thus affirming their least or most important value). A score of 0 was given when participants chose the filler answer. Thus, total scores on the affirmation task ranged from 0 to 10, with higher scores indicating higher affirmation. We expected that self-affirmed participants would endorse the manipulated value more often than non-affirmed participants.

Implicit affect. Participants were asked to complete a seemingly unrelated study on the esthetic evaluation of words in which they evaluated the pleasantness of some non-words on a 7-point scale anchored at unpleasant-pleasant (cf. Jostmann, Koole, Van der Wulp, & Fockenberg, 2005; Tesser, Millar, & Moore, 1988). They were told that non-words (“pleban”, “lempon”, “tokitorer”, Stapel, 2002) were chosen to avoid disturbing influences of word meaning and were instructed to simply report their initial reactions. The average rating given on this pleasantness scale served as our measure of implicit affect (α = .60), with higher scores indicating more positive affect.

Accessibility of threat-related cognitions. As in Study 4.1, participants completed a similar lexical decision task to assess the accessibility of threat-related cognitions. In total, participants responded to 64 words, consisting of 32 existing words and 32 non-words. Among the existing words, 10 words were threat-related (e.g., toxic, cancer, disease), the remaining items reflected control words (e.g., table, car, building). All words were matched
on word length across word type categories, randomly presented, and preceded by four practice trials.

Results

Manipulation Checks

Self-threat level. A one-way ANOVA showed that participants in the high self-threat condition felt significantly more vulnerable to the health problems ($M = 4.33, SD = 1.31$) than participants in the moderate self-threat condition ($M = 2.51, SD = 0.97$), $F(1, 77) = 48.68, p < .001, \eta_p^2 = .39$.4

Self-affirmation. A one-way ANOVA confirmed that self-affirmed participants endorsed the manipulated value more often ($M = 6.56, SD = 2.02$) than non-affirmed participants ($M = 3.30, SD = 1.60$), $F(1, 77) = 63.28, p < .001, \eta_p^2 = .45$.

Implicit Affect

A 2 (self-threat level) x 2 (self-affirmation status) ANOVA on implicit affect revealed no significant effects (all $p$s > .60).5

Accessibility of Threat-Related Cognitions

The computer recorded both the response (i.e., word or non-word) and response time (in milliseconds) per presented item for each participant. Incorrect responses were excluded from the analysis (2.89% across the experimental trials, which were evenly distributed across conditions). The mean RT for threat-related words and control words were subjected to a 2 (self-threat level: moderate vs. high) x 2 (self-affirmation status: non-affirmed vs. self-affirmed) x 2 (word type: threat-related vs. control) mixed design ANOVA with repeated measures on the third factor. As expected, the analysis revealed an interaction effect between self-threat level, self-affirmation, and word type, $F(1, 75) = 7.44, p = .008, \eta_p^2 = .09$. Simple effects analysis indicated that in the moderate self-threat condition, self-affirmed participants
responded faster to threat-related words than non-affirmed participants, $F(1, 75) = 4.44, p = .039, \eta_p^2 = .06$. This effect of self-affirmation was reversed in the high self-threat condition. Self-affirmed participants responded slower to threat-related words than non-affirmed participants, $F(1, 75) = 5.16, p = .026, \eta_p^2 = .06$. These effects did not occur for the responses to the control words (moderate self-threat condition, $F[1, 75] = 2.05, p = .16$; high self-threat condition, $F < 1$). Relevant means are displayed in Table 4.3.

Table 4.3. Mean RT in milliseconds as a function of self-threat level, self-affirmation status condition, and word type (Study 4.2)

<table>
<thead>
<tr>
<th></th>
<th>Moderate self-threat</th>
<th>High self-threat</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Non-affirmed</td>
<td>Self-affirmed</td>
</tr>
<tr>
<td>Threat-related</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-affirmed</td>
<td>606$_a$</td>
<td>546$_b$</td>
</tr>
<tr>
<td>Self-affirmed</td>
<td>598$_a$</td>
<td>555$_a$</td>
</tr>
<tr>
<td>Control</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Means with different subscripts (in rows, for each self-threat condition) differ significantly from each other ($p < .05$).

Discussion

The results of Study 4.2 demonstrate that the positive effect of self-affirmation on people’s implicit responsiveness to threatening health information is limited to conditions in which people experience a moderate threat to the self. Specifically, self-affirmation increased the accessibility of threat-related cognitions under conditions of moderate self-threat. In contrast, and in accordance with our prediction, self-affirmation decreased the accessibility of threat-related cognitions under conditions of high self-threat. These results conceptually replicate earlier findings showing that self-threat level moderates the impact of self-affirmation on the (explicit) processing of threatening health information (Van Koningsbruggen & Das, 2008b). Thus, the current findings demonstrate the limits of self-
affirmation beyond the domain of explicit cognitive responses. Because information that is less accessible will be less likely encoded and recalled (Carlston & Smith, 1996; Higgins, 1996), these findings also suggest that self-affirmed people are less likely to integrate threatening information into the self-system in situations in which they experience a high threat to the self. Moreover, since self-affirmation did not influence implicit affect, it seems unlikely that affect accounts for the present findings.

**General Discussion**

The present research provided a first test of the impact of self-affirmation on people’s implicit responsiveness to threatening health information. We predicted that when self-affirmation diminishes concerns about self-integrity, threatening health information would become easier to retrieve from memory. However, in line with recent findings (Van Koningsbruggen & Das, 2008b), we expected that this beneficial effect of self-affirmation would be limited to conditions in which people experience a moderate threat to the self.

Consistent with this idea, Study 4.1 demonstrated that self-affirmation increased accessibility of threat-related cognitions under moderate self-threat conditions. Additionally, Study 4.1 showed that self-affirmation promoted perceptions of message quality and intentions to take precautions when moderately threatened. Study 4.2 demonstrated that the positive effect of self-affirmation on people’s implicit responsiveness to threatening health information is indeed limited to moderately threatening conditions. The results of Study 4.2 showed that self-affirmation increased accessibility of threat-related cognitions under moderate self-threat conditions (cf. Study 4.1), but decreased accessibility of threat-related cognitions under high self-threat conditions. The present findings underscore that self-affirmation may promote or impede responsiveness to threatening health information depending on the level of self-threat (cf. Van Koningsbruggen & Das, 2008b), hence
demonstrating the limits of self-affirmation beyond the domain of explicit cognitive responses. In addition, Study 4.2 ruled (implicit) affect out as an alternative explanation for the observed effects as self-affirmation did not influence affect in this study.

Notwithstanding the fact that self-affirmation may backfire under high self-threat conditions, Study 4.1 shows that self-affirmation can reduce defensiveness toward threatening health information, and can promote adaptive behavioral intentions among people for whom the information is relevant (cf. Harris & Napper, 2005; Sherman et al., 2000). Additionally, these findings are the first to demonstrate that the beneficial effects of self-affirmation are not limited to changes in explicit variables related to persuasive outcomes but extend to the realm of less controlled, automatic cognitive responses to threatening health information. Although in Study 4.1 the accessibility of threat-related cognitions did not mediate the effects of self-affirmation on perceptions of message quality and intentions among participants in the moderate self-threat condition, we did find a significant relationship between the accessibility of threat-related cognitions and perceived message quality among self-affirmed participants in the moderate self-threat condition. This finding shows that implicit responsiveness to threatening health information is indeed related to decreased defensive responding at the explicit level, at least under conditions that are moderately threatening to the self.

In addition, this finding is in line with previous research showing that global cognitions may be related to specific cognitions particularly when they are highly accessible (Fazio, 1995). In Study 4.1, it is likely that we could not establish mediation because we measured construct accessibility rather than attitude accessibility (Fazio, 1995; see also, e.g., Houston & Fazio, 1989), the former being a less likely candidate in predicting specific attitudes. Furthermore, some features of the Study 4.1 seem to have limited our ability to detect these effects; we assessed perceived message quality with only one item, and our data may have limited statistical power to detect mediation (Fritz & MacKinnon, 2007). Future
studies should take these limitations into account when further examining the relation between accessibility of threat-related cognitions and variables involved in the behavior change process. This would also provide a better test of whether the change in implicit responsiveness observed in Study 4.1 is indicative of a mechanism underlying the obtained self-affirmation effects on the explicit measures, or indicative of a consequence of more open-minded information processing fostered by self-affirmation.

The present research also confirmed the importance of carefully attending to self-threat level in determining whether self-affirmation will have beneficial effects or not on health information processing. When moderately threatened, self-affirmation seems to diminish self-integrity concerns and to stimulate integration of the threat into the self-system (cf. Sherman & Cohen, 2006), enabling people to encode threatening aspects of the health information because these aspects no longer threaten people’s self-integrity. When severely threatened, however, self-affirmation seems to reinforce concerns over self-integrity rather than diminish them (cf. Van Koningsbruggen & Das, 2008b). The findings of Study 4.2 show that under these conditions, self-affirmation seems to prevent integration of the threat into the self-system, thereby impeding people to encode threatening aspects of the health information. Whether the decreased implicit responsiveness observed in Study 4.2 backfires on explicit responses to health risk information should be more specifically addressed in future research.

In conclusion, this research provided a first test of the impact of self-affirmation on implicit responsiveness to threatening health information. While several studies demonstrated how self-affirmation affects responses to threatening health information at the explicit level, insight into the cognitive processes instigated by self-affirmation tended to be limited (Sherman & Cohen, 2006). The present findings represent a step forward in this area: self-affirmation promotes or impedes accessibility of threatening health information, depending on the level of self-threat people experience when faced with such information.