

Chapter 6

Reconciling Self-Protection with Self-Improvement

Self-Affirmation Theory

DAVID K. SHERMAN
KIMBERLY A. HARTSON

A persistent puzzle emerges from a review of the psychological research literature on how people defend the self from potentially threatening information and events: People can be adamantly protective of their sense of self-worth and defensively resistant to threatening information, on the one hand, and yet capable of accepting threatening, critical information and changing their behavior, on the other. The general tendency for self-enhancement that is the focus of so many of the chapters in this book (see also Sedikides & Gregg, 2008; Taylor & Brown, 1988) is a testament to people's ability to maintain a view of the self as capable, adaptive, and culturally appropriate despite the many threats that they may encounter. However, only a very maladaptive organism would be so engaged in self-protection that it was incapable of change. And so the question is raised as to how individuals can put aside their ego-protective needs when faced with criticism and threats and come to recognize that potentially painful and self-threatening pursuits could also help one improve.

For example, at times people will resist information from their doctors that they need to change their health behavior, from their teachers that they need to improve their study skills, and from their family members that they need to adjust their interpersonal style. Yet, despite the self-threat that comes with acknowledging imperfections, flaws, and mistakes, individuals can come to change such important behaviors. In this chapter, we propose that one way that people can resolve the tension between self-protection and self-improvement is by affirming the self in an important domain of self-worth in response to threat, thereby reaffirming an overall image of self-integrity. Since the inception of self-affirmation theory

(Steele, 1988), there have been many demonstrations, in both the laboratory and the field, that when people affirm the self, they are less defensive and more open to otherwise threatening information (see Sherman & Cohen, 2006, for a review). Studies that have used self-affirmation theory with individuals confronting threats have also found that affirmations, such as having people write about their important values, can reduce physiological stress responses (Sherman, Bunyan, Creswell, & Jaremka, 2009) and can improve academic performance among those experiencing high levels of stress due to stereotype threat (Cohen, Garcia, Apfel, & Master, 2006; Cohen, Garcia, Purdie-Vaughns, Apfel, & Brzustoski, 2009). How is it that simply reminding oneself of one's important values can exert such effects? In this chapter, we propose a multistage model drawing on recent research that suggests that self-affirmations may reduce threat, stress, and defensiveness by boosting self-resources (Schmeichel & Vohs, 2009) and changing one's perspective on threats (Critcher & Dunning, 2009; Crocker, Niiya, & Mischkowski, 2008; Sherman, Cohen, et al., 2009; Wakslak & Trope, 2009).

Overview and Goals for This Chapter

Self-affirmation theory begins with the premise that people are motivated to maintain the perceived worth and integrity of the self and examines how people respond to information and events that threaten a valued self-image (Steele, 1988; see also Aronson, Cohen, & Nail, 1999; McQueen & Klein, 2006; Sherman & Cohen, 2006). Everyday life offers numerous potential psychological threats, whether it be poor performance on an exam or health information implying that past behaviors put one at risk of disease. People often respond to such threats in a manner that leads them to construe situations as less threatening to personal worth and well-being.

For example, people respond to failure at times by attributing it externally (Miller & Ross, 1975) or by simply disidentifying with the threatened domain, sustaining self-worth but forestalling self-improvement (Major, Spencer, Schmader, Wolfe, & Crocker, 1998; Nussbaum & Steele, 2007). People view contradictory information through the prism of their ideology, placing greater scrutiny on information inconsistent with prior beliefs (Lord, Ross, & Lepper, 1979). When health information suggests personal risk, people may react by challenging the information rather than by changing their risky behavior (Kunda, 1987). These defensive judgments help maintain the perceived integrity of the self by reducing potential psychological threats.

The most basic tenet of self-affirmation theory is that people are motivated to protect the perceived integrity and worth of the self (Sherman & Cohen, 2006; Steele, 1988). People seek to maintain a *global* sense of self-integrity, “a phenomenal experience of the self . . . as adaptively and morally adequate, that is, competent, good, coherent, unitary, stable, capable of free choice, capable of controlling important outcomes” (Steele, 1988, p. 262) rather than just of their perceived worth in a specific domain or in particular situations. This global self-integrity is best thought of not as self-esteem or a positive feeling toward the self but as a quality possessed by the “self-system.” This self-system (see Figure 6.1) is composed of the different domains that are important to an individual, including individuals' roles (e.g., being a teacher or a sibling); values (e.g., being charitable or moral); social identities (e.g., as a member of a particular group); and belief systems (e.g., political parties or religion;

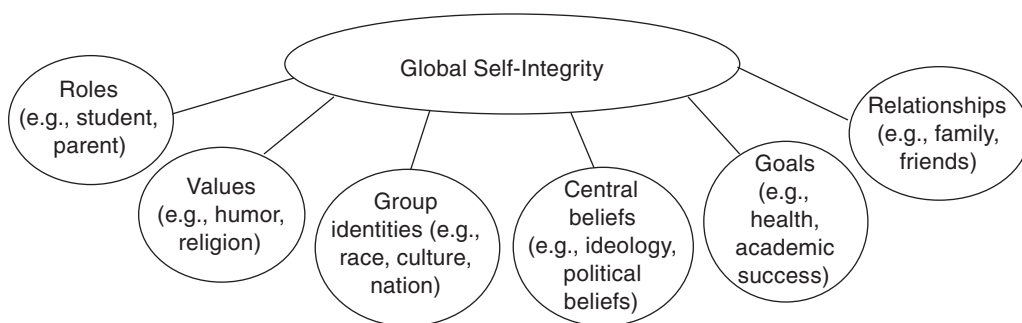


FIGURE 6.1. Schematic representation of self-system.

Sherman & Cohen, 2006). The goal of the self-system is to maintain this quality of global self-integrity. When this goal of global self-integrity maintenance is threatened, people seek means of reaffirming their self-integrity.

To borrow a familiar metaphor, the self-system could be thought of as an individual's self-concept (i.e., the iceberg), with the working self-concept (Markus & Nurius, 1986) being whatever roles, values, or identities are salient for the individual at the moment (i.e., the tip of the iceberg). Within this flexible self-system, all of these domains or self-conceptions are potential routes to global self-integrity that could be either threatened or affirmed.

Thus defensive responses to threatening events and information (e.g., rationalizing one's risk for health problems as minimal due to one's limiting of smoking) are one means of affirming global self-integrity, because they reduce the psychological threat inherent in the information. Similarly, changing one's behavior can also minimize the threat (e.g., quitting smoking). However, the insight of the model is that the self-system is flexible, so if people can affirm an important domain of self-worth in another domain, they will have less need to rationalize away threatening information because their overall self-integrity, their view of the self as being capable and adaptive, has been secured (Sherman & Cohen, 2006; Steele, 1988). Indeed, many studies have demonstrated that experimentally induced self-affirmations—for example, writing about important values—can reduce defensive processing of threatening information (Adams, Tormala, & O'Brien, 2006; Binning, Sherman, Cohen, & Heitland, in press; Cohen, Aronson, & Steele, 2000; Fein & Spencer, 1997; Jaremka, Bunyan, Collins, & Sherman, in press; Monin, Sawyer, & Marquez, 2008; Unzueta & Lowery, 2008). These affirmations of self-integrity can serve as reminders that self-worth is not solely contingent on the threatened domain, reducing the need to defend the self in the particular threatened domain.

For the remainder of this chapter, we would like to accomplish three primary goals. First, to provide a background for the theory, we situate self-affirmation as a process that operates as part of a psychological immune system (Gilbert, Pinel, Wilson, Blumberg, & Wheatley, 1998) that is engaged when individuals experience self-threats. Second, we review recent research on the effects of self-affirmation on individuals confronting various threats to self-integrity stemming from health threats, stressful situations, and environments of stereotype threat. Third, we present a model that lays out how self-affirmation manipulations exert their effects.

The Psychological Immune System: Insights on Affirmation and Self-Protection

Self-affirmation may be considered as one process that operates as part of a psychological immune system that is engaged when individuals experience self-threats (Gilbert et al., 1998). Gilbert and colleagues introduced this concept of the psychological immune system, including self-affirmation among many self-enhancement strategies:

Psychologists from Freud to Festinger have described the artful methods by which the human mind ignores, augments, transforms, and rearranges information in its unending battle against the affective consequences of negative events (Festinger, 1957; Freud, 1936; Greenwald, 1980; Kunda, 1990; Steele, 1988 ... Taylor & Brown, 1988). Some of these methods are quite simple ... and some are more complicated ... taken in sum, however, they seem to constitute a *psychological immune system* that serves to protect the individual from an overdose of gloom. (Gilbert et al., 1998, p. 619)

Just as the body's immune system responds to pathogens and protects against disease by identifying and killing foreign invaders and tumor cells, the psychological immune system initiates protective adaptations under impending threats to the self. And just as the actual immune system has different components to protect the body from illness, such as lymph nodes and the spleen, the psychological immune system has different strategies that it employs to protect the self from threat. Motivated inferences, self-serving judgments, rationalizations, self-enhancement, positive illusions, and self-affirmation are all processes that are helpful in protecting the self. Additional components of the psychological immune system that initiate protective responses to threat include active coping, seeking social support, and increasing one's effort to nullify the threat. Each process is likely to have its own particular function, advantages, and disadvantages, and yet, overall, they work together as an integrated system. Thus the first insight we can take from the psychological immune system metaphor is to consider self-affirmation as one strategy among many that people employ with the goal of maintaining psychological health and to examine how it achieves this goal.

The second insight derived from this metaphor is the notion that there are costs and benefits to engaging the threat-response system. That is, a response of the physical immune system (e.g., a fever that kills an invading agent) requires a great mobilization of energy that alters equilibrium, detracts from long-term projects (such as growth and reproduction), and leaves the body more prone to attacks from other pathogens. Similarly, engaging the psychological immune system is not without costs. Motivated inferences, defensive attributions, and self-enhancement are all psychological responses that could, at times, lead individuals to ignore potentially important information, especially if these responses are chronically engaged.

However, there are other costs that we would like to emphasize, specific costs to the individual that are a direct consequence of the threat response. Consider a woman who is preparing to take a math test that could potentially confirm a negative stereotype about her gender—that is, the type of situation that prompts stereotype threat (Steele, 1997, 2010). People respond to stereotype threat with a physiological stress response, a tendency to actively monitor performance (i.e., to be overly focused on the threat), and an effort to suppress negative thoughts and feelings (Schmader, Johns, & Forbes, 2008). These costs may

be the by-product of processes that yield psychological benefits, such as attempts to actively disconfirm the negative stereotype (Jamieson & Harkins, 2007). Yet these various responses consume cognitive resources needed to perform well on the test (Logel et al., 2009). Furthermore, the desire to disprove the stereotype may prevent the relaxed looseness needed to excel on high-stakes tests, as people spend more time toiling on individual problems rather than working efficiently on the test as a whole (Steele & Aronson, 1995).

Thus it may be that certain psychological responses to threatening situations—such as an active attempt to disconfirm the stereotype or to actively monitor performance—may paradoxically lead to the eventual outcome—underperformance—that a person most wants to avoid (see also Wegner, 1997, who makes a related point about suppression). We would like to suggest, however, that other responses of the psychological immune system could potentially offset these costs and enable people to deal with threats without some of the negative consequences just noted. Indeed, recent studies demonstrating that self-affirmation can reduce stress (Creswell et al., 2005; Sherman, Bunyan, et al., 2009) and underperformance in situations of stereotype threat (Cohen et al., 2006; Martens, Johns, Greenberg, & Schimel, 2006) suggest that this particular aspect of the psychological immune system, when induced experimentally, could buffer individuals from some of the pernicious effects of threatening environments.

The third insight stemming from the psychological immune system is the view that it may operate more effectively when people are unaware of its operation. Gilbert et al. (1998) argue that people are generally unaware of the influence of the psychological immune system, a phenomenon they term *immune neglect*. Research suggests that immune neglect characterizes self-affirmation as well (Sherman, Cohen, et al., 2009). People are generally unaware of the impact of self-affirmation, and when awareness of the influence of an affirmation is increased, effects are attenuated. This insight has implications for understanding self-affirmation effects, as well as for efforts to apply the theory in intervention settings, a point we shall return to later in this chapter.

Research Review: Recent Research Using Self-Affirmation Theory

We next review recent research on the effects of self-affirmation on individuals confronting various threats to self-integrity. In particular, we focus on three different threat responses: (1) defensiveness in response to threatening health information; (2) physiological stress responses to threatening situations; and (3) underperformance in situations of social identity threat. We focus on these three areas because they combine both laboratory and field studies and are relevant to important real-world outcomes such as health, stress, and grades. We also focus on these three areas because, although the effects are conceptually similar and theoretically consistent, the domains of application are sufficiently diverse as to raise important questions as to how reflecting on values and personal characteristics—the common self-affirmation manipulation employed across the studies—can yield consistent effects.

Reducing Defensiveness to Threatening Health Information

Health information from one's doctor, the media, or health campaigns has the potential to threaten an individual's self-perception of being healthful, smart, and adaptive by suggesting

that a person has acted unwisely or participated in unhealthy, risky behaviors. Consequently, individuals may respond to these messages defensively, maintaining their positive perceptions of the self by disregarding or downplaying the risks described in the health message (Kunda, 1987). For example, a smoker confronted with the evidence that smoking leads to cancer may question the evidence or downplay the extent to which he or she is at risk, thinking "I only smoke a cigarette a day." These defensive responses are exhibited most strongly among those for whom the message is most relevant (Lieberman & Chaiken, 1992), paradoxically leading individuals who are at the highest risk away from adaptive behavioral change.

To accept threatening health messages requires acknowledging the maladaptiveness of one's own behavior. Self-affirmation provides a psychological buffer for individuals to accept these health messages without sacrificing their self-perceptions as globally competent and worthy individuals. When people write about important values, for example, they are reminded of what makes them competent and worthy in a global sense, and thus threatening health information can be evaluated in terms of its importance for personal health and not its implications for self-worth. Across a variety of domains, research has found that experimentally induced self-affirmations can reduce defensiveness and increase acceptance of health messages (see Harris & Epton, 2009, for review).

For example, one study on smoking behavior exposed university student smokers to disturbing images that were going to be printed on actual packs of cigarettes in the European Union (Harris, Mayle, Mabbott, and Napper, 2007). These images graphically depicted the consequences of prolonged smoking (e.g., open heart surgery) and were the latest in a trend of fear-inducing antismoking campaigns. The study was designed to test the effectiveness of these images and similar threat-driven campaigns at deterring their target audience, smokers, from smoking.

After completing a self-affirmation task (writing about one's desirable features) or a control task (writing about what one recently ate), participants rated the graphic smoking images for how threatening, unpleasant, and personally relevant they found them and then completed a number of questionnaires about their current smoking behavior, intentions to quit smoking, and thoughts and feelings related to smoking.

The researchers found that participants in the no-affirmation condition defensively reduced the potential threat in this information by rating the graphic images as less threatening and self-relevant. By contrast, participants who were affirmed were more able to acknowledge the potential threat inherent in the information and see it as relevant (Harris et al., 2007). Additionally, affirmed participants expressed stronger intentions to change their smoking behavior (which persisted for 1 week), more negative feelings toward smoking, and more feelings of control over their smoking behavior than nonaffirmed participants. However, self-affirmed participants did not differ from nonaffirmed participants in actual smoking behavior in the week following the affirmation.

In another study conducted with heavy smokers at a factory in the United Kingdom, researchers presented smokers with a leaflet adapted from the U.K. government's antismoking campaign (Armitage, Harris, Hepton, & Napper, 2008). Participants who completed a self-affirmation by indicating agreement with statements related to kindness had greater acceptance of the antismoking information, increased intentions to reduce their smoking behavior, and were more likely to take a brochure with further tips on how to quit smoking relative to participants in a noaffirmation control condition.

The effects of self-affirmation on health message acceptance have been shown to persist

for time periods ranging up to 1 month. In one study (Harris & Napper, 2005), women who ranged in their alcohol consumption from light to heavy drinkers read a pamphlet linking alcohol to breast cancer. Participants wrote about important values in the self-affirmation condition or relatively unimportant values and why they might be important to others in the no-affirmation condition. Perceived risk of developing breast cancer was assessed immediately following their reading of the pamphlet, 1 week later, and 1 month later. Among the heavy drinkers, who were the most at risk and therefore most threatened by the pamphlet, the women who were affirmed saw themselves as more at risk for breast cancer, reported being more able to imagine themselves with breast cancer, and had stronger intentions to change their alcohol consumption than those who were not affirmed, and these perceptions persisted through the 1-month follow-up. However, despite intentions to change, high-risk self-affirmed participants did not demonstrate a reduction in alcohol consumption in the month following the study (Harris & Napper, 2005). One possible mechanism consistent with these findings is that the affirmations enable people to orient more toward the threatening information. Recent research using an implicit-attention paradigm demonstrates that moderately heavy drinkers who are affirmed are biased toward threatening words linking alcohol use and breast cancer (Klein & Harris, 2009; see also van Koningsbruggen, Das, & Roskos-Ewoldsen, 2009), whereas nonaffirmed participants who are moderately heavy drinkers showed a bias away from such threatening words.

Thus self-affirmation has been shown to be effective at leading individuals to take the first behavioral steps toward changing their health behaviors, such as forming intentions (Harris & Napper, 2005) and taking relevant brochures (Armitage et al., 2008). Consistent findings were also obtained in a study with sexually active college students, who watched a video that was meant to imply that the students' current behavior put them at risk for potentially contracting sexually transmitted diseases such as HIV (Sherman, Nelson, & Steele, 2000). Those in the affirmation condition increased their perceived risk of acquiring HIV following the video, whereas those in the no-affirmation control condition responded to the video by maintaining their pretest levels of perceived risk. Moreover, following the video, when participants were given the opportunity to purchase condoms, 50% of affirmed participants did so, whereas only 25% of nonaffirmed participants did.

However, taking brochures (Armitage et al., 2008) and purchasing condoms (Sherman et al., 2000) are still only proxies for the critical behavioral outcome of long-term and enduring change; evidence for this kind of change in the health context has been weak (Harris & Epton, 2009; McQueen & Klein, 2006). Moreover, effects of self-affirmation have not been observed in all health studies, as some studies have found null effects of the affirmation manipulations on health-risk acceptance (Dillard, McCaul, & Magnan, 2005) and intentions to adopt detection behaviors (Fry & Prentice-Dunn, 2005). Thus it is an open question as to when, or whether, self-affirmation can lead to actual health behavior change.

One study explored whether self-affirmation would lead to health behavioral change when the change was focused more on health promotion (e.g., eating healthier foods) than on terminating unhealthy behaviors (e.g., quitting smoking; Epton & Harris, 2008). Female participants completed either a self-affirmation task or a control task prior to reading a health message advocating the beneficial health effects of eating fruits and vegetables. Participants in the self-affirmation condition not only intended to consume more fruits and vegetables immediately following reading the article but also actually consumed 5.5 more portions of fruits and vegetables during the week immediately following the experimental session (Epton

& Harris, 2008). Thus aspects of the behavior (e.g., promoting health vs. preventing illness) may moderate when the affirmations lead to successful behavior change.

In the health domain, self-affirmation has thus been shown to decrease defensiveness to health messages, increase risk perceptions and intentions to change behavior, and in some cases even facilitate behavior change. Enabling people to affirm values and other important aspects of the self can lead them to accept the maladaptiveness of their current health behaviors and instigate adaptive behavioral change.

Reducing Physiological Responses to Stressful Situations

Students often get “stressed out” about exams, just as employees get stressed out about performance evaluations and patients get stressed out about medical tests. One reason these events are stressful is that they have implications for how the student, the employee, and the patient see themselves in important domains. Naturalistic stressors of this sort are a common feature of day-to-day life and are known to trigger a cascade of neuroendocrine events that are adaptive in the short run but that can, over time, lead to increased susceptibility to negative mental and physical health outcomes (Cohen, Janicki-Deverts, & Miller, 2007). Identifying psychological means by which individuals can cope adaptively with stressful situations is a topic of historical and contemporary research interest (Carver, 2007; Miller & Cohen, 2001). Recently, researchers have sought to use self-affirmation theory to understand why people may experience stress and to incorporate self-affirmation interventions into stress-reduction techniques.

The self-affirmation analysis of stress posits that affirming valued sources of self-worth such as important personal qualities, values, or relationships can buffer threats to the self, reducing the impact that these threats have on both physiological and psychological responses (Sherman & Cohen, 2006). Because people are motivated to maintain a *global* sense of self-integrity, rather than their perceived worth in a specific domain or in particular situations, affirmations of unrelated domains of self-worth may make self-evaluation less contingent on a particular focal stressor. As a consequence, stressors may be experienced as less taxing.

Evidence for this approach has been obtained in studies in which participants completed self-affirmation (or control) tasks during or prior to experiencing a stressful event. One study investigated the effect of self-affirmation on stress reduction in response to acute laboratory stressors (Creswell et al., 2005). Following either an affirmation (indicating agreement with important personal values on a scale) or a control (indicating agreement with unimportant values) task, participants were put through the Trier Social Stress Task (Kirschbaum, Pirke, & Hellhammer, 1993), a performance situation in which they had to give a 5-minute speech describing why they were qualified for a job as an administrative assistant in the psychology department and then had to count backward aloud from 2,083 by 13's.

Cortisol, an indicator of hypothalamic–pituitary–adrenocortical (HPA) activation, was assessed via saliva samples taken throughout the session. During baseline, there were no differences in cortisol levels between the affirmation group and the control group. Twenty minutes after the stress task, however, participants in the control group had elevated stress hormone cortisol levels, whereas those in the affirmation condition showed no change in cortisol levels from baseline. This pattern of stress reactivity persisted for 45 minutes follow-

ing the stress task, providing evidence that self-affirmation can buffer individuals from stress associated with threatening environments.

To examine whether self-affirmation could buffer individuals when confronting more chronic, naturalistic stressors, a study was conducted with students at the time of their most stressful midterm examinations (Sherman, Bunyan, et al., 2009). By allowing students to affirm other valued aspects of the self, self-affirmation might provide a buffer from the threat to the self posed by the demanding school and test-taking environment. Participants indicated what midterm examination they were most stressed about and provided urine samples collected during two 15-hour intervals 14 days prior to the exam (baseline) and on the morning of the exam. Catecholamine levels were assessed from the urine samples. Catecholamines, specifically epinephrine and norepinephrine, are released by the sympathetic nervous system in order to mobilize energy for the fight-or-flight response to a stressor (Lundberg, 2000).

In the 2 weeks prior to their exams, the students wrote two essays as part of an online writing exercise either on the values most important to them (self-affirmation condition) or on values unimportant to them (control condition). Whereas participants in the control condition experienced an increase in epinephrine levels from baseline to the morning of the exam, individuals in the self-affirmation condition had epinephrine levels that did not differ significantly from baseline (Sherman, Bunyan, et al., 2009). Further, the affirmation seemed to be the most beneficial for those who were the most threatened and therefore most stressed, as effects were strongest among those who, at baseline, expressed the most concern about negative academic performance evaluations.

As stress increases one's susceptibility to colds and serious illnesses (Cohen, Tyrrell, & Smith, 1993; Marmot, Bosma, Hemingway, Brunner, & Stansfeld, 1997) an important question is whether the stress-buffering effects of self-affirmation could yield health benefits. To test the hypothesis that self-affirmation could buffer individuals from the negative health outcomes associated with daily stressors, researchers in one study had undergraduates write essays over winter break about the events of the day and how they related to their most important value (affirmation condition) or on a number of control topics (Keough & Markus, 1999). Those in the affirmation condition reported being less stressed and visited the health center less often than those in control conditions (conditions in which people wrote about what happened that day, about positive things that happened that day, or did not write about anything; Keough & Markus, 1999). Thus repeated experimentally induced self-affirmations can reduce stress-induced health symptoms.

Finally, initial evidence for the beneficial effects of self-affirmation among ill populations comes from a study with early-stage breast cancer survivors (Creswell et al., 2007) who had participated in a study on expressive writing in which they were assigned to write essays on different topics related to their cancer (their deepest thoughts and feelings regarding the disease, the benefits they found since being diagnosed, or the facts of the day). The original study by Stanton and colleagues (2002) found reduced symptoms and doctor visits in the 3 months following the study for women who wrote about their thoughts and feelings related to cancer and those in the emotional-processing and benefit-finding conditions. In the later study (Creswell et al., 2007), all essays were coded for evidence of self-affirmation, operationalized as positive reflections on valued domains. The amount of self-affirming writing predicted reductions in distress and physical symptoms among the breast cancer survivors and, in fact, explained much of the positive effects of expressive writing on health outcomes.

In sum, self-affirmation appears to be able to reduce stress in chronically ill populations, highlighting the potential for greater application of self-affirmation interventions in health and medical contexts.

Attenuating the Effects of Stereotype Threat on Performance

Academic environments can be particularly threatening, as intellectual evaluation is a persistent element in school and doing well is important to how many individuals see themselves. Stereotype threat is an additional burden that occurs for individuals when a negative stereotype is made salient about their group's performance within a certain domain (Steele & Aronson, 1995). In academic settings, in which stereotypes about race and intellectual ability or gender and quantitative ability are pervasive, stereotype threat can lead to underperformance and disidentification (Steele, 1997).

In both the laboratory and the field, several researchers have examined whether self-affirmation can help individuals from negatively stereotyped groups overcome stereotype threat and improve their performance in situations in which they have previously been shown to underperform. One series of studies examined whether self-affirmation would lead to improved performance by women on a math test (Martens et al., 2006). It was theorized that, by allowing females to affirm another valued domain, self-affirmation would reduce the threat women face of confirming the stereotypes associated with math performance. Thus self-affirmation directly addresses the threat and stress response posed by the stereotype, unlike other interventions aimed at reducing stereotype threat that seek to undermine the stereotype itself by portraying a test as gender-fair (Spencer, Steele, & Quinn, 1999) or making an alternative identity salient (Rydell, McConnell, & Beilock, 2009).

In one study, male and female undergraduates completed either an affirmation task or a control task prior to taking a difficult math test (Martens et al., 2006). The test was described as either diagnostic of their math abilities, making it stereotype-threat inducing, or as a test that was under development for research purposes, making it unrelated to math ability and therefore not stereotype-threat inducing. Women in the stereotype-threat condition performed worse than women in the no-stereotype-threat condition and worse than men in the stereotype-threat condition. However, women in the stereotype-threat condition who completed the self-affirmation performed significantly better than women in the stereotype-threat condition without affirmation. A second study found that affirmation reduced the sex difference in spatial rotation under conditions of stereotype threat (Martens et al., 2006). Thus affirmation buffered female students from the threat associated with confirming negative stereotypes.

The findings in these laboratory studies led to important questions regarding self-affirmation as an intervention to be used in classroom settings to reduce the psychological distress associated with stereotype threat and to potentially improve minority students' academic performance. A series of field studies (Cohen et al., 2006, 2009) investigated the effect that affirming important aspects of the self could have on easing the evaluative stress that minority group members feel when faced with the threat of confirming negative stereotypes about their racial group.

Cohen and colleagues initially looked at self-affirmation's effects on the academic performance of two different student cohorts over the course of one school term (Cohen et al., 2006). During the beginning of the seventh-grade school year, students were given the experi-

mental task once or twice during the term as part of their regular class curriculum. The task consisted of a 15-minute standardized writing exercise in which students in the affirmation condition were asked to write about values that they had indicated as highly important to them, whereas those in the control condition wrote about values that they had indicated as unimportant.

The fall term grades for all students revealed that African American students in the affirmation condition earned higher grades in the course in which the affirmation was given than African American students in the control condition, and this effect was largest for those who had initially performed the worst (Cohen et al., 2006). That is, the affirmation was most effective for those African American students who had the most room for improvement. By contrast, the affirmation had no effect on European American students' academic performance. Based on a comparison between African American and European American students' performance across conditions, the self-affirmation intervention reduced 40% of the racial achievement gap that had existed between these students prior to the intervention (Cohen et al., 2006).

Cohen and colleagues (2009) conducted a follow-up study to assess the longer term impact of the self-affirmation intervention—that is, whether affirming values could yield academic benefits for the 2-year period following the affirmation. During the course of the seventh-grade year, the experimental task was given to each of three cohorts three to five times at approximately equal intervals. The researchers monitored academic performance over the course of the 2-year period. They found that the grade point averages of African American students were higher for those in the affirmation condition than for those in the control condition, even over 2 years. Consistent with previous findings (Cohen et al., 2006), the self-affirmation intervention was most effective for those whose performance started out the worst; by contrast, African American students in the no-affirmation condition who had low initial performance did not improve over time. Analysis of the individual tests of students suggests that the affirmation did not so much improve grades as it prevented grades from dropping; analogously, in the stress study reported earlier (Sherman, Bunyan, et al., 2009), the affirmation did not reduce epinephrine levels but kept levels from increasing during the stressful examination time.

In academic settings, individuals face consistent threats to the self, which can be magnified for those in stereotyped groups. Across a variety of studies in both the field (Cohen et al., 2006, 2009) and the laboratory (Martens et al., 2006), writing about values seems to buffer students from the evaluative stress associated with stereotype threat, leading to benefits in academic performance.

Summary

Self-affirmation exerted consistent effects across the disparate domains of responses to personally relevant health information, physiological responses to stressful situations, and performance among individuals contending with a negative group stereotype. This particular aspect of the psychological immune system, the reaffirmation of self-integrity by reflecting on alternative domains of self-worth, seems to enable people to respond to threatening events and information with less defensiveness, keeping stress at bay, and preventing decrements in performance. Important questions to address are why, how, and when self-affirmations exert such effects.

Understanding Self-Affirmation Effects: A Multistage Approach

We believe that self-affirmation effects may be understood in terms of the following three stages:

1. Affirmation boosts self-resources—the psychological resources that one has to cope with a threat.
2. With self-resources salient, an individual can view the threat from a broader perspective.
3. This broader perspective allows the threat and the self to become “uncoupled,” reducing the threat’s potency at affecting the self.

We elaborate and present evidence supporting these three steps in the sections to follow.

Affirmation Boosts Self-Resources

An influential model of how to conceptualize psychological resources in stressful or threatening situations was advanced by Hobfoll (1989). He defined resources as “those objects, personal characteristics, conditions, or energies that are valued by the individual or that serve as a means for attainment of these objects, personal characteristics, and energies” (p. 516). Psychological resources are likely to be taxed when individuals experience a focal threat or stressor, but they can also encompass aspects of the self that are unrelated to the threat. For example, one of the most commonly chosen affirmation topics is relations with friends and family (Creswell et al., 2007; Crocker et al., 2008). Writing about personal relationships or the other values within an affirmation activity could equip people with additional psychological resources when they experience threats, potentially leading to the observed effects previously reviewed.

However, do self-affirmation manipulations boost self-resources? It certainly seems plausible, considering the basic values affirmation that is most commonly employed as a self-affirmation manipulation (McQueen & Klein, 2006). People write about, or indicate the importance of, their most important values via a values scale, values that generally are unrelated to the threatening event. These manipulations are typically ideographic in that people select and write about values or personal characteristics that are important to them; for more general affirmations (e.g., of one’s kindness; Armitage et al., 2008) to be effective, the domain would have to be important to many people. Note also that writing about failure in an important domain would not be self-affirming, and indeed such topics have been used as a control condition in some studies, yielding different effects from standard affirmations (e.g., greater, rather than reduced, closed-mindedness; Cohen et al., 2007). The intent of affirmation manipulations is to make individuals aware of important aspects of the self so that they can consider the potential threat within the context of their overall self-image. By writing about religion, or friends and family, or other values, the individual is reminded of core aspects of life and resources that can be relied on when confronting threats.

Yet, to date, the evidence that the self-affirmation manipulation boosted self-resources in this manner was rather indirect, demonstrating predicted effects on measures theorized to be reflective of having greater resources (i.e., reduced stress and defensiveness). However, recent studies by Schmeichel and Vohs (2009) present more direct evidence that affirmations

can boost self-resources. Research on self-control has shown quite dramatically that acts of self-control deplete a resource required for subsequent self-control (Muraven, Tice, & Baumeister, 1998). When resources are depleted, for example, by being forced to delay gratification or to maintain a straight face during an emotional video, individuals perform worse at subsequent tasks that require self-control (Muraven & Baumeister, 2000).

If self-affirmation boosts self-resources, then it should be able to counteract this type of ego depletion; and, if this were the case, it would suggest that the self-resources that researchers in self-affirmation have written about (Steele, Spencer, & Lynch, 1993) possess similar qualities to the self-resources that ego-depletion researchers have studied (Baumeister, Bratslavsky, Muraven, & Tice, 1998). That hypothesis was supported across several studies (Schmeichel & Vohs, 2009). For example, in one study, some participants' self-resources were depleted by being instructed to inhibit use of the letter *n* in writing a story, whereas other participants could write without such prohibitions. Then participants wrote about either an important value or a relatively unimportant value, the self-affirmation manipulation. It was theorized that writing about important values would make salient additional self-resources that a person could draw on during subsequent self-control tasks. Finally, researchers measured how long participants could keep their hands in cold water, the cold-pressor task that is used as a measure of self-control. In the no-affirmation condition, the standard effects of resource depletion were exhibited, as those who depleted resources by regulating their writing kept their hands in the water for less time than those who could write freely. This difference was eliminated among those in the self-affirmation condition, as those who were self-affirmed after resource depletion performed just as well as those not depleted.

This interaction effect between self-affirmation and ego depletion was replicated in another study using a different manipulation of resource depletion (watching a video while *not* attending to words written on the screen) and a different measure of self-control (persistence on puzzles), supporting the generality of the findings (Schmeichel & Vohs, 2009, Study 2). This set of studies presents the strongest evidence to date that self-affirmation boosts self-resources, enabling people to confront ego-depleting, threatening events without typical threat responses. Considering these findings in terms of the reviewed studies on self-affirmation, then, it may be that threatening health information, stressful situations, and stereotype threat all consume psychological resources and lead to ego depletion, but that affirmation counteracts negative outcomes by providing additional self-resources with which people can confront threats.

With Boosted Self-Resources, an Individual Can View the Threat from a Broader Perspective

The next question to address is how a boost in self-resources might reduce typical threat responses, as exhibited in self-affirmation studies. We propose that the extra self-resources that the affirmation makes salient may change the way an individual perceives the threat.¹ That is, an individual under threat may perceive that his or her entire self-evaluation is con-

¹One question that may have emerged from the previous section is whether, given the effects of glucose on ego depletion (Gailliot & Baumeister, 2007), self-affirmation effects can be reduced to increased energy or even glucose levels. Although it seems that glucose can boost self-control, it seems doubtful that it changes one's perspective on a threat, although, of course, this remains an open question for future research.

tingent on the threatened domain, but a threatened individual whose self-resources were just boosted via an affirmation may take a broader view of the threat. Consider the study described earlier (Sherman, Bunyan, et al., 2009) in which participants who affirmed important values showed reduced sympathetic nervous system responses on the day of their most stressful exam relative to nonaffirmed participants. After taking the exam, participants responded to questions assessing their appraisals of the exam, such as: “During the exam I often thought about what would happen if I failed.” Self-affirmed participants reported reduced stress appraisals on such items, suggesting that the self-affirmation may have reduced their meta-evaluative concerns, a finding consistent with research demonstrating that self-affirmation can lead to the cessation of rumination after failure (Koole, Smeets, van Knippenberg, & Dijksterhuis, 1999).

These findings suggest that when people are given the opportunity to write about important values, people may be more secure in their self-worth and thus less concerned about what the potential failure would represent in terms of their overall self-image, enabling them to “transcend concerns about self-image or self-worth” (Crocker et al., 2008, p. 741) by, for example, focusing more on others than on themselves (Crocker et al., 2008). However, this evidence that affirmation changes the individual’s perspective is still somewhat indirect and inferential. More direct evidence that affirmation can change people’s perspective comes from Schmeichel and Vohs (2009) and Wakslak and Trope (1999), who both theorized that self-affirmations enable people to view events at higher levels of abstraction, or what has been termed higher levels of construal (Trope & Liberman, 2003; Vallacher & Wegner, 1989). Levels of construal have been associated with self-control, suggesting that, at the dispositional level, those who see things at lower levels are more driven by impulsive responses to situational threats and less able to forestall the negative long-term consequences of their behavior (Vallacher & Wegner, 1989). By contrast, those who view things at a higher level are more able to integrate actions and information with their values (Vallacher & Wegner, 1989).

Self-affirmation manipulations typically focus people on important (vs. unimportant) values. These values are abstract and lead people to think about their ideals at a relatively high level, to examine why a particular value is important to the self. Thus they may lead individuals to see things from a higher level of construal. This hypothesis was tested in two different studies (Schmeichel & Vohs, 2009; Wakslak & Trope, 2009), in which participants completed a standard self-affirmation manipulation and then the Behavioral Identification Form (Vallacher & Wegner, 1989), a standardized measure that presents 25 behaviors and asks participants to describe whether a given behavior (e.g., “taking a test”) is best seen as a low level of construal (“answering questions”) or a higher level of construal (“showing one’s knowledge”). Self-affirmation led participants to view events at a higher level of construal (Schmeichel & Vohs, 2009; Wakslak & Trope, 2009). Moreover, self-affirmations led people to evaluate items in terms of their broader structure, rather than focusing on their secondary details (Wakslak & Trope, 2009). Finally, self-affirmations that were induced at a higher level of construal (e.g., writing about *why* one pursues a particular value) led to greater self-control than self-affirmations that were induced at a lower level of construal (e.g., writing about *how* one pursues a particular value; Schmeichel & Vohs, 2009).

Thus affirmations appear to lead people to take a broader view of events in general. However, the question remains: Do they enable people to see *threatening* events in particular with greater perspective? Critcher and Dunning (2009) propose that affirmations expand the

working self-concept by reminding people that the threatened domain does not encompass the entire self. Values affirmations remind people of other aspects of the self-concept and thus reduce the implications that the threatened domain has for self-evaluation, making defensive biases unnecessary. As a result, as Sherman and Cohen (2006) suggested:

When global perceptions of self-integrity are affirmed, otherwise threatening events or information lose their self-threatening capacity because the individual can view them within a broader, larger view of the self. People can thus focus not on the implications for self-integrity of a given threat or stressor, but on its informational value. When self-affirmed, individuals feel as though the task of proving their worth, both to themselves and to others, is “settled.” As a consequence, they can focus on other salient demands in the situation beyond ego protection. (p. 189)

Several recent studies support this notion. First, after writing about an important identity unrelated to academics (e.g., being a cultured individual), college student participants were asked to think about their intended majors (Critcher & Dunning, 2009). Affirmed students were more likely to agree with statements such as, “In thinking of domains that contribute to how I feel about myself, nonacademic aspects easily come to mind” and “There is a lot more to my skills and abilities than just who I am in my academic major” than nonaffirmed students. Thus thinking about a valued identity seemed to enable students to consider academics—the source of much potential threat to their self-concept—as one part of many and as less central to how they see themselves. Importantly, the affirmation did not lead students to trivialize academics, consistent with research showing that self-affirmation does not lead people to think a threatened domain is less important (Correll, Spencer, & Zanna, 2004; see also Brinol, Petty, Gallardo, & DeMarree, 2007). Rather, those who completed a self-affirmation were more certain of who they were, and thus their self-evaluations were less contingent on the threatened domain. And, indeed, self-affirmations have been shown to increase self-concept clarity, suggesting that people are more certain about who they are when given the chance to write about important values (Wakslak & Trope, 2009). When these findings are taken together, it appears that self-affirmations can reduce the ego-defensive needs prompted by threat by enabling people to view threats within a broader context of the self.

When Affirmed, the Threat and the Self Can Become “Uncoupled,” Reducing the Threat’s Potency at Affecting the Self

As summarized previously, there is evidence that self-affirmations can boost self-resources, changing people’s construal of events, which enables them to view the threat within the broader perspective of the self. We propose one further consequence of this broader perspective that links the proposed mechanism to the outcomes of interest: Self-affirmation may reduce defensiveness by “uncoupling” the threatened domain from self-evaluation. Operationally, this process could yield weaker correlations between measures of self-evaluation and measures related to the threatened domain in self-affirmation conditions, relative to no-affirmation conditions, with strong correlations indicating that the individual is anchoring social judgments on the self (Dunning, 2003).

Several studies have found this pattern of correlations, suggesting the uncoupling effect of self-affirmation (Cohen et al., 2007; Sherman, Cohen, et al., 2009; Sherman & Kim,

2005). For example, in one field study, intramural athletes who participated in a competitive team sports event were assigned to the affirmation or no-affirmation condition and then indicated their attributions for their team's winning or losing. In the absence of self-affirmation, self-attributions were highly, positively correlated with group attributions, suggesting that people were anchoring their judgments of the group on the self (Sherman & Kim, 2005). However, those who had completed a self-affirmation task no longer used the self as an anchor in judging their group. That is, their team assessments were no longer correlated with self-perceptions, suggesting that the athletes evaluated the group and the self independently of each other.

Similar patterns have been observed in evaluations of identity-threatening information. In one study, participants' patriotism was negatively correlated with openness of information critical of United States foreign policy, with patriots being less open to this information than antipatriots (people who described themselves as critics of the United States) (Cohen et al., 2007). However, when their national identity was made salient, self-affirmation attenuated this relationship, as participants evaluated the information critical of the United States independently of their personal feelings toward the country. Consequently, patriots were more open, and antipatriots less accepting, of the information. Similar findings were obtained in another study (Sherman, Cohen, et al., 2009), in which, in the absence of affirmation, identification with a particular sports team (the San Francisco Giants) negatively predicted how people responded to information critical of a member of that team (Barry Bonds), such that highly identified fans were the least open to the potentially, identity threatening information. By contrast, when participants completed a standard self-affirmation, there was no correlation between identification and evaluation of the information, as the highly identified Giants fans become more open to the information critical of Barry Bonds.

In the domain of performance under social identity threat, further evidence for uncoupling comes from affirmation intervention studies with middle school children (Cohen et al., 2009). For affirmed minority students, performing poorly early in the school year did not have any bearing on their sense of adequacy in school at the end of the year. But for nonaffirmed minority students, it did, as early poor performance was related to lower feelings of self-adequacy at the end of the academic year. Affirmation severed the relationship between a given threat (poor academic performance) and long-term self-evaluations, here over the course of a year.

Within the domain of evaluation of health information, evidence for uncoupling comes from studies where, in the absence of affirmation, there exists a negative relationship between a personal characteristic, such as the relevance of the health message, and evaluation of the potentially threatening information. For example, in one study (Sherman et al., 2000), in the absence of affirmation, coffee drinkers were more resistant than non-coffee-drinkers to health information describing the negative effects of caffeine. By contrast, when affirmed, this negative relationship was reversed, and the coffee drinkers were most open to the potentially threatening health information (see also Harris & Napper, 2005). As these findings suggest, self-affirmation may in some cases reverse correlational relationships (suggesting not just uncoupling, but reverse coupling) and as other researchers have suggested, may at times decrease positive relationships (Klein & Monin, 2009). For an extended and detailed discussion of the relationships among variables in regard to self-affirmation and health information, see also Harris and Epton (2010) and Klein and Monin (2009). As both the present discussion of uncoupling and these other publications point out, understanding how

self-affirmation affects the relationship among variables may yield important information in understanding how, when, and why affirmation manipulations exert their effects.

To summarize, recent studies conducted across a number of different laboratories have provided evidence for a multistage process by which value affirmations buffer the self and reduce stress and defensiveness and improve performance. A story suggests itself. Affirmations boost individuals' self-resources by reminding them of other aspects of the self not centrally relevant to the threat. In so doing, affirmations broaden an individual's perspective on the threat, enabling them to view potential threats at a higher level of construal. With this broader perspective, people are able to evaluate threats to a greater extent, independently of ego-defensive concerns. Although we do not claim that this is "the underlying process" of all self-affirmation effects, given the multitude of studies that have been conducted across many domains, we believe that significant progress has now been made toward understanding why self-affirmation manipulations yield their effects. Future research should build on these initial findings by conducting studies that directly link these proposed mechanisms to outcomes such as defensiveness, stress, and underperformance.

We address two final issues that we believe are important to understanding self-affirmation's effects. The first issue addresses how brief self-affirmation manipulations, such as writing about values, could result in long-term behavior change (Cohen et al., 2009; Harris & Napper, 2005). The second issue addresses boundary conditions for self-affirmation effects, focusing on the role of awareness in the affirmation process (Sherman, Cohen, et al., 2009).

Recursive Processes in Self-Affirmation

Self-affirmation manipulations have had effects over 1 month on health intentions (Harris & Napper, 2005), 1 week on eating behaviors (Epton & Harris, 2008), 10 days on evaluations of President Obama among Republicans (Binning et al., in press) several weeks on sympathetic nervous system activation (Sherman, Bunyan, et al., 2009), and up to two years on academic performance (Cohen et al., 2009). How might this occur? At a general level, the affirmed state—the thoughts and feelings one has after writing about values—is likely to be relatively brief. However, the change in how individuals view and respond to threatening events may be more likely to persist over longer periods of time.

Cohen and colleagues (Cohen et al., 2009) have focused on the recursive nature of self-affirmation processes. In a context of persistent threats, such as those that occur for minority students in academic settings, they propose that:

A recursive cycle, where psychological threat lowers performance, increasing threat and lowering performance further, in a repeating process, can magnify early performance differences among students. Early outcomes set the starting point and initial trajectory of a recursive cycle and so can have disproportional influence. For instance, the low self-confidence of students who experience early failure, even by chance, is surprisingly difficult to undo. A well-timed intervention could provide appreciable long-term performance benefits through early interruption of a recursive cycle. (Cohen et al., 2009; p. 400)

Their 2-year follow-up of a series of self-affirmation interventions conducted in mixed-race middle schools provides striking evidence that affirmation interrupted the recursive pro-

cess whereby difficulty begets failure. Consider three findings in particular. First, although overall the affirmation produced significantly improved grades for African-American students, the effects were strongest for initially low-achieving African Americans. That is, prior performance was less predictive of postintervention performance (suggesting uncoupling). Second, in this same study, the affirmation reduced the slope of the downward trend in performance for threatened students. Relative to nonaffirmed students, who experienced a steep decline term by term, affirmed African American students maintained their performance.

Third, as noted before, in the absence of affirmation, early poorly performing African American students perceived themselves to be less adequate and reported that they fit in less well at school at the end of the academic year relative to higher performing students. That is, their self-evaluations were tightly linked to their performance in the threatened domain. By contrast, affirmed participants exhibited a decoupling between their perceptions of adequacy in school and their earlier performance. Thus lower performing African American students had greater perceptions of personal adequacy when affirmed than not. In sum, among minority students, early poor performance instigated a downward slide in performance and self-perceived adequacy. This downward slide was halted in the affirmation condition. Beyond interrupting a recursive cycle, the affirmation seemed to cause enduring changes in how people construed poor performance. They saw poor performance as having fewer implications for their self-perceived adequacy.

Boundary Conditions for Self-Affirmation Effects

Given the self-affirmation findings described in this chapter that have shown beneficial effects on important outcomes, there may be a sense that implementing self-affirmation widely, broadly, and frequently could yield positive effects. There are two points that we would like to make in this regard. First, the “beneficial” effects of self-affirmations are typically a function of the construction of the study (and likely to reflect the underlying interests of the researcher in increasing acceptance of health information or attenuating stress and stereotype threat, etc.) and not necessarily a function of the theory or how the self-system functions. The more general point made by the self-affirmation studies reviewed in this paper is that self-affirmations can reduce the need to defend a particular threatened identity or aspect of the self. However, under the right conditions, it is likely that self-affirmations could lead to “harmful” effects. Indeed, when an open-minded identity was made salient, self-affirmation reduced the need to defend that identity, and people were more closed-minded in response to threatening information (Cohen et al., 2007, Study 3). Thus, although studies conducted with self-affirmation typically yield beneficial effects, this need not be the case. To the extent that defensive responses result in “positive” outcomes, such as a person who argues against a racist attack because it threatens his or her egalitarian worldview, it would be predicted that self-affirmations would attenuate this (positive) defensive response as well.

The second point refers to how the affirmations are implemented. Although affirmation interventions may produce large effects due to the recursive nature of the threat–affirmation process, we suggest that they should also adhere to small-scale subtlety (Sherman, Cohen, et al., 2009). The self-affirmation process is not a panacea for stress, threat, and defensiveness

but, rather, as we outlined earlier, plays a role in the operation of the “psychological immune system” that people use to protect the self when it is threatened (Gilbert et al., 1998). In keeping with that reasoning, recent work has focused on a relevant feature of the psychological immune system and its implication for self-affirmation; namely, that people are generally unaware of it when it is operating. Heightened awareness of self-affirmation processes may attenuate the effects described earlier.

In particular, three facets of awareness of the affirmation process have been explored (Sherman, Cohen, et al., 2009). First, people can be affirmed without deliberative awareness. Affirmations that are manipulated via value-relevant sentence-unscrambling procedures and are hence outside of participants’ awareness can exhibit similar threat-reducing effects to those of standard affirmation manipulations. When participants are queried after affirmation studies about factors that may have influenced them, they do not spontaneously generate anything related to the affirmation manipulation as a potential cause of their judgments about the threatening information, and, when asked, people rate affirmation tasks (e.g., writing tasks) as relatively minor influences on their judgments or performance. Finally, when participants are aware of the effects of affirmation, as assessed by measure (those who spontaneously report that they were influenced by the affirmation task) or manipulation (via studies in which participants are informed of potential affirmation effects), affirmation effects are weaker (Sherman, Cohen, et al., 2009).

Increased awareness may attenuate affirmation effectiveness through several of the processes described earlier. First, the boost in resources that occurs when a person writes about important values may be reduced to the extent to which people are focused on the outcome of the affirming event. When people explicitly pursue happiness, or explicitly strive to boost their self-esteem, such acts can be self-defeating and lead to less happiness and reduced self-esteem, because they may lead people to focus more on the extrinsic benefits of an act (Crocker & Park, 2004; Schooler, Ariely, & Loewenstein, 2003). Consistent with this notion, affirmations that focus on intrinsic aspects of the self are more successful at reducing defensiveness than affirmations that focus on extrinsic aspects of the self (Schimel, Arndt, Banko, & Cook, 2004).

Second, heightened awareness of an affirmation in the face of threat may lead people to link the affirmation to the threatened domain rather than broadening their perspective on the threat (Critcher & Dunning, 2009). If people perceive that they are engaged in a stress-reduction exercise, for example, they may be more cognizant of their stressors rather than of the alternative self-resources that are made salient by the affirmation. This issue may also speak to the types of affirmations that can backfire, or lead to greater defensiveness. Research has found that same-domain affirmations exacerbate dissonance, whereas alternative domain affirmations reduce it (Blanton, Cooper, Skurnik, & Aronson, 1997; Sivanathan, Molden, Galinsky, & Ku, 2008). It may be that same-domain affirmations cause people to link the affirmation exercise with threat, which may lead it to backfire.

These findings on awareness have implications for future field research, as they offer a perhaps counterintuitive suggestion for those interested in applying self-affirmations in field settings. The key to an effective affirmation intervention may lie in the subtlety of its delivery and the minimalism of its administration. It may be that more transparent affirmations may raise awareness and reduce effectiveness.

Conclusion: Reconciling Self-Protection with Self-Improvement

We began this chapter by raising the puzzle of how the tendency for self-enhancement and self-protection that is so prevalent and well documented seems to imply that people could never accept self-criticism and make positive behavioral changes. Based on the research reviewed here, self-affirmation appears to be one mechanism in the psychological immune system that helps explain this paradox. People are willing to be self-critical when they feel globally self-affirmed. Self-affirmation can thus lead to self-improvement in terms of less defensiveness and stress and more positive behavioral change and better performance. Moreover, the process underlying self-affirmation effects is beginning to come into focus. Values affirmations appear to boost self-resources, broadening the perspective with which people view threats, and enabling them to reconcile protection and self-critical motivations.

Acknowledgments

We acknowledge the support of National Science Foundation Grant No. 0720429. We also thank Mark Alicke, Kevin Binning, Cameron Brick, Thai Chu, Geoffrey Cohen, P. J. Henry, Lisa Jar-emka, and Nate Way for commenting on earlier versions of this chapter, and Peter Harris and William Klein for their discussion of the issues raised in this chapter.

References

- Adams, G., Tormala, T. T., & O'Brien, L. T. (2006). The effect of self-affirmation on perception of racism. *Journal of Experimental Social Psychology*, 42, 616–626.
- 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., Cohen, G. L., & Nail, P. R. (1999). Self-affirmation theory: An update and appraisal. In E. Harmon-Jones & J. Mills (Eds.), *Cognitive dissonance theory: Revival with revisions and controversies* (pp. 127–147). Washington, DC: American Psychological Association.
- Baumeister, R. F., Bratslavsky, E., Muraven, M., & Tice, D. M. (1998). Ego depletion: Is the active self a limited resource? *Journal of Personality and Social Psychology*, 74, 1252–1265.
- Binning, K. R., Sherman, D. K., Cohen, G. L., & Heitland, K. (in press). Seeing the other side: Reducing political partisanship via self-affirmation in the 2008 presidential election. *Analyses of Social Issues and Public Policy*.
- 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.
- Brinol, P., Petty, R. E., Gallardo, I., & DeMarree, K. G. (2007). The effect of self-affirmation in non-threatening persuasion domains: Timing affects the process. *Personality and Social Psychology Bulletin*, 33, 1533–1546.
- Carver, C. (2007). Stress, coping, and health. In H. S. Friedman & R. C. Silver (Eds.), *Foundations of health psychology* (pp. 117–144). New York: Oxford University Press.
- Cohen, G. L., Aronson, J., & Steele, C. M. (2000). When beliefs yield to evidence: Reducing biased evaluation by affirming the self. *Personality and Social Psychology Bulletin*, 26, 1151–1164.
- Cohen, G. L., Garcia, J., Apfel, N., & Master, A. (2006). Reducing the racial achievement gap: A social-psychological intervention. *Science*, 313, 1307–1310.

- Cohen, G. L., Garcia, J., Purdie-Vaughns, V., Apfel, N., & Brzustoski, P. (2009). Recursive processes in self-affirmation: Intervening to close the minority achievement gap. *Science*, 324, 400–403.
- Cohen, G. L., Sherman, D. K., Bastardi, A., McGoey, M., Hsu, A., & Ross, L. (2007). Bridging the partisan divide: Self-affirmation reduces ideological closed-mindedness and inflexibility in negotiation. *Journal of Personality and Social Psychology*, 93, 415–430.
- Cohen, S., Janicki-Deverts, D., & Miller, G. E. (2007). Psychological stress and disease. *Journal of the American Medical Association*, 298, 1685–1687.
- Cohen, S., Tyrrell, D. A. J., & Smith, A. P. (1993). Negative life events, perceived stress, negative affect, and susceptibility to the common cold. *Journal of Personality and Social Psychology*, 64, 131–140.
- 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.
- Creswell, J. D., Lam, S., Stanton, A. L., Taylor, S. E., Bower, J. E., & Sherman, D. K. (2007). Does self-affirmation, cognitive processing, or discovery of meaning explain cancer-related health benefits of expressive writing? *Personality and Social Psychology Bulletin*, 33, 238–250.
- Creswell, J. D., Welch, W., Taylor, S. E., Sherman, D. K., Gruenewald, T., & Mann, T. (2005). Affirmation of personal values buffers neuroendocrine and psychological stress responses. *Psychological Science*, 16, 846–851.
- Critcher, C. R., & Dunning, D. (2009). *Affirmations provide perspective: Reducing defensiveness by expanding the working self-concept*. Manuscript in preparation.
- 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.
- Crocker, J., & Park, L. E. (2004). The costly pursuit of self-esteem. *Psychological Bulletin*, 130, 392–414.
- Dillard, A. J., McCaul, K. D., & Magnan, R. E. (2005). Why is such a smart person like you smoking? Using self-affirmation to reduce defensiveness to cigarette warning labels. *Journal of Applied Biobehavioral Research*, 10, 165–182.
- 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.), *Motivated social perception: The Ontario symposium* (Vol. 9, pp. 45–72). Mahwah, NJ: Erlbaum.
- Epton, T., & Harris, P. R. (2008). Self-affirmation promotes health behavior change. *Health Psychology*, 27, 746–752.
- Fein, S., & Spencer, S. J. (1997). Prejudice as self-image maintenance: Affirming the self through derogating others. *Journal of Personality and Social Psychology*, 73, 31–44.
- Festinger, L. A. (1957). *A theory of cognitive dissonance*. Evanston, IL: Row, Peterson.
- Freud, A. (1936). *Ego and the mechanisms of defense* (Rev. ed., C. Baines, Trans.). New York: International Universities Press.
- Fry, R. B., & Prentice-Dunn, S. (2005). Effects of coping information and value affirmation on responses to a perceived health threat. *Health Communication*, 17, 133–147.
- Gailliot, M. T., & Baumeister, R. F. (2007). The physiology of willpower: Linking blood glucose to self-control. *Personality and Social Psychology Review*, 11, 303–327.
- Gilbert, D. T., Pinel, E. C., Wilson, T. D., Blumberg, S. J., & Wheatley, T. P. (1998). Immune neglect: A source of durability bias in affective forecasting. *Journal of Personality and Social Psychology*, 75, 617–638.
- Greenwald, A. G. (1980). The totalitarian ego: Fabrication and revision of personal history. *American Psychologist*, 35, 603–618.
- Harris, P. R., & Epton, T. (2009). The impact of self-affirmation on health cognition, health behaviour and other health related responses: A narrative review. *Social and Personality Psychology Compass*, 3, 962–978.
- Harris, P. R., & Epton, T. (2010). The impact of self-affirmation on health-related cognition and health behavior: Issues and prospects. *Social and Personality Psychology Compass*, 4, 439–454.

- 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.
- Hobfoll, S. E. (1989). Conservation of resources: A new attempt at conceptualizing stress. *American Psychologist, 44*, 513–524.
- Jamieson, J. P., & Harkins, S. G. (2007). Mere effort and stereotype threat performance effects. *Journal of Personality and Social Psychology, 93*, 544–564.
- Jaremka, L. M., Bunyan, D. P., Collins, N. L., & Sherman, D. K. (in press). Reducing defensive distancing: Self-affirmation and risk regulation in response to relationship threats. *Journal of Experimental Social Psychology*.
- Keough, K. A., & Markus, H. R. (1999). On being well: The role of the self in building the bridge from philosophy to biology. *Psychological Inquiry, 9*, 49–53.
- Kirschbaum, C., Pirke, K., & Hellhammer, D. H. (1993). The "Trier Social Stress Test": A tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology, 28*, 76–81.
- Klein, W. M. P., & Harris, P. R. (2009). Self-affirmation enhances attentional bias toward threatening components of a persuasive message. *Psychological Science, 20*, 1463–1467.
- 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.
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin, 108*, 480–498.
- Liberman, A., & Chaiken, S. (1992). Defensive processing of personally relevant health messages. *Personality and Social Psychology Bulletin, 18*, 669–679.
- Logel, C., Walton, G. M., Spencer, S. J., Iserman, E. C., von Hippel, W., & Bell, A. (2009). Interacting with sexist men triggers social identity threat among female engineers. *Journal of Personality and Social Psychology, 96*, 1089–1103.
- Lord, C., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology, 37*, 2098–2109.
- Lundberg, U. (2000). Catecholamines. In G. Fink (Ed.), *Encyclopedia of stress* (Vol. 1, pp. 408–413). San Diego, CA: Academic Press.
- Major, B., Spencer, S. J., Schmader, T., Wolfe, C., & Crocker, J. (1998). Coping with negative stereotypes about intellectual performance: The role of psychological disengagement. *Personality and Social Psychology Bulletin, 24*, 34–50.
- Markus, H., & Nurius, P. (1986). Possible selves. *American Psychologist, 41*, 954–969.
- Marmot, M., Bosma, H., Hemingway, H., Brunner, E., & Stansfeld, S. (1997). Contribution of job control and other risk factors to social variations in coronary heart disease. *Lancet, 350*, 235–239.
- Martens, A., Johns, M., Greenberg, J., & Schimel, J. (2006). Combating stereotype threat: The effect of self-affirmation on women's intellectual performance. *Journal of Experimental Social Psychology, 42*, 236–243.
- McQueen, A., & Klein, W. (2006). Experimental manipulations of self-affirmation: A systematic review. *Self and Identity, 5*, 289–354.
- Miller, D. T., & Ross, M. (1975). Self-serving biases in the attribution of causality: Fact or fiction? *Psychological Bulletin, 82*, 213–225.
- Miller, G. E., & Cohen, S. (2001). Psychological interventions and the immune system: A meta-analytic review and critique. *Health Psychology, 20*, 47–63.
- Monin, B., Sawyer, P. J., & Marquez, M. J. (2008). The rejection of moral rebels: Resenting those who do the right thing. *Journal of Personality and Social Psychology, 95*, 76–93.

- Muraven, M., & Baumeister, R. F. (2000). Self-regulation and depletion of limited resources: Does self-control resemble a muscle? *Psychological Bulletin*, 126, 247–259.
- Muraven, M., Tice, D. M., & Baumeister, R. F. (1998). Self-control as a limited resource: Regulatory depletion patterns. *Journal of Personality and Social Psychology*, 74, 774–789.
- Nussbaum, A. D., & Steele, C. M. (2007). Situational disengagement and persistence in the face of adversity. *Journal of Experimental Social Psychology*, 43, 127–143.
- Rydell, R. J., McConnell, A. R., & Beilock, S. L. (2009). Multiple social identities and stereotype threat: Imbalance, accessibility, and working memory. *Journal of Personality and Social Psychology*, 96, 949–966.
- Schimmel, J., Arndt, J., Banko, K. M., & Cook, A. (2004). Not all self-affirmations were created equal: The cognitive and social benefits of affirming the intrinsic (vs. extrinsic) self. *Social Cognition*, 22, 75–99.
- Schmader, T., Johns, M., & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological Review*, 115, 336–356.
- Schmeichel, B. J., & Vohs, K. D. (2009). Self-affirmation and self-control: Affirming core values counteracts ego depletion. *Journal of Personality and Social Psychology*, 96, 770–782.
- Schooler, J. W., Ariely, D., & Loewenstein, G. (2003). The pursuit and assessment of happiness can be self-defeating. In I. Brocas & J. Carrillo (Eds.), *The psychology of economics decisions: Vol. 1. Rationality and well-being* (pp. 41–70). Oxford, UK: Oxford University Press.
- Sedikides, C., & Gregg, A. P. (2008). Self-enhancement: Food for thought. *Perspectives on Psychological Science*, 3, 102–116.
- 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., Bunyan, D. P., Creswell, J. D., & Jaremka, L. M. (2009). Psychological vulnerability and stress: The effects of self-affirmation on sympathetic nervous system responses to naturalistic stressors. *Health Psychology*, 28, 554–562.
- 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 the process of self-affirmation. *Journal of Personality and Social Psychology*, 97, 745–764.
- Sherman, D. K., & Kim, H. S. (2005). Is there an “I” in “team”? The role of the self in group-serving judgments. *Journal of Personality and Social Psychology*, 88, 108–120.
- Sivanathan, 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., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women’s math performance. *Journal of Experimental Social Psychology*, 35, 4–28.
- Stanton, A. L., Danoff-Burg, S., Sworowski, L. A., Collins, C. A., Branstetter, A., Rodriguez-Hanley, A., et al. (2002). Randomized, controlled trial of written emotional expression and benefit finding in breast cancer patients. *Journal of Clinical Oncology*, 20, 4160–4168.
- 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.
- Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, 52, 613–629.
- Steele, C. M. (2010). *Whistling Vivaldi and other clues to how stereotypes affect us*. New York: Norton.
- Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, 69, 797–811.

- Steele, C. M., Spencer, S. J., & Lynch, M. (1993). Self-image resilience and dissonance: The role of affirmational resources. *Journal of Personality and Social Psychology*, 64, 885–896.
- Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin*, 103, 193–210.
- Trope, Y., & Liberman, N. (2003). Temporal construal. *Psychological Review*, 110, 403–421.
- Unzueta, M. M., & Lowery, B. S. (2008). Defining racism safely: The role of self-image maintenance on white Americans' conceptions of racism. *Journal of Experimental Social Psychology*, 44, 1491–1497.
- Vallacher, R. R., & Wegner, D. M. (1989). Levels of personal agency: Individual variation in action identification. *Journal of Personality and Social Psychology*, 57, 660–671.
- van Koningsbruggen, G., Das, E., & Roskos-Ewoldsen, D. R. (2009). How self-affirmation reduces defensive processing of threatening health information: Evidence at the implicit level. *Health Psychology*, 28, 563–568.
- Wakslak, C., & Trope, Y. (2009). Cognitive consequences of affirming the self: The relationship between self-affirmation and object construal. *Journal of Experimental Social Psychology*, 45, 927–932.
- Wegner, D. M. (1997). When the antidote is the poison: Ironical mental control processes. *Psychological Science*, 8, 148–150.