Imagine a student who receives negative feedback on his exam performance. He feels threatened by this negative result but feels better after denigrating a member of a minority group (Fein & Spencer, 1997). However, when reminded of success in another part of life, he seems to be able to tolerate this threat without engaging in prejudice toward other people. Now, imagine a White American who is reminded of the negative historical treatment inflicted by her fellow White Americans on African Americans. This citizen feels threatened by this information and denies the impact of racism on other people. An opportunity to affirm the self by writing about core values, however, leads to greater acknowledgment of wrongdoing by the in-group (Adams, Tormala, & O'Brien, 2006).

According to self-affirmation theory (Cohen & Sherman, 2014; Steele, 1988), people are motivated to maintain self-integrity, an image of themselves as capable and adaptive. Receiving negative information that threatens the individual or the collective self—as illustrated by the examples above—can lead people to display more prejudicial judgments toward others as a way of protecting self-integrity (Sherman & Cohen, 2006). Negative intergroup attitudes may be particularly effective in restoring self-integrity (Fein & Spencer, 1997) by providing an opportunity for downward social comparison (Fein, Hoshino-Browne, Davies, & Spencer, 2003). Furthermore, out-groups may provide a potential explanation (a “scapegoat”) for negative events that happen to the self (Rothschild, Landau, Sullivan, & Keefer, 2012).

The link between self-integrity and intergroup attitudes was first examined in a series of studies in which individuals were threatened with negative feedback about their individual self (i.e., personal identity; Fein & Spencer, 1997). Receiving self-image-threatening information led participants to evaluate out-group members more stereotypically; however, affirming participants’ core values eroded the link between self-image threats and expressions of prejudice.

Since the seminal research by Fein and Spencer (1997), an important development has emerged: an examination of the threat to the collective level of the self (i.e., social identity) that people face when negative
Self-Affirmation and Prejudice Reduction

acts are committed by their in-group. According to social-identity theory (Tajfel & Turner, 1979), to the extent that individuals derive feelings of self-esteem from their in-group, to recognize negative actions and transgressions committed by members of one's in-group (e.g., genocide against ethnic minorities) can be self-threatening. To reduce that threat, people may deny their in-group's responsibility and instead blame the out-group to restore self-worth.

In this article, we focus on empirical studies that show how social-identity threat can be buffered by self-affirming, which in turn can lead to less-negative defensive reactions (e.g., prejudice toward out-groups). While many affirmation studies are targeted toward in-group threats (i.e., wrongful in-group actions), less research is focused on perceived out-group threats and the prejudice resulting from such threats. We review research showing that both types of threat can be buffered by self-affirmation. Affirmation can also target the individual (self-affirmation) as well as the collective self (group affirmation). However, self- and group affirmation are not equally effective in reducing prejudice, and in this article, we explain why. We also examine for whom self-affirmations are likely to be most successful, by taking into account the role of individual differences such as cultural value orientations (e.g., individualistic vs. collectivistic).

We present a model wherein the effect of affirmation on the reduction of defensive reactions is mediated by the perception of threats (Fig 1). The effectiveness of affirmations, we propose, is further moderated by both individual differences and cultural norms. Although both factors potentially moderate either self-affirmation or group affirmation, the extant research suggests more strongly that the impact of self-affirmation depends on individual differences, whereas the impact of group affirmation can vary as a function of the cultural norms promoted in the context in which the affirmation is implemented. This model, developed from results of existing studies, is proposed as a guide for future research to better investigate when and why affirmations can lead to prejudice reduction.

In-Group Versus Out-Group Threats

Considering that self-affirmation theory (Sherman & Cohen, 2006; Steele, 1988) was developed to clarify how people cope with threats, we begin our review by focusing on the different types of social-identity threats people face in intergroup contexts (the theorized mediator in Fig 1). Internal in-group threats (the theorized mediator in Fig 1). Internal in-group threats may stem from negative actions taken by the in-group (Rothgerber, 1997; see Table 1 for examples). Individual defensive responses to internal in-group threats, such as denying collective responsibility for wrongdoing, can reduce one's threat perception and repair a threatened self-image.

Given that intergroup conflict is steeped in historical context, in-group threat can occur when people reflect on the historical origins of a threat and attempt to parse responsibility. After September 11, 2001, American participants evaluated an article that criticized the role of U.S. foreign policy in fomenting conditions that led to the attack (Cohen et al., 2007). The more patriotic people were, the more negatively they evaluated the information; however, highly patriotic Americans who received a self-affirmation were more open to the information criticizing the United States. Similarly, in Bosnia and Herzegovina, self-affirmation increased Serbian participants' willingness to acknowledge in-group responsibility for the Srebrenica genocide, to express feelings of group-based guilt, and to support reparation policies (Čehajić-Clancy, Effron, Halperin, Liberman, &

Fig. 1. Theoretical model of the impact of self- and group affirmation on defensive reactions resulting from in-group and out-group threats, as moderated by individual differences and cultural norms. Vertical arrows indicate whether the strength of the moderators' influence is high (solid arrows) or low (dashed arrows).
Ross, 2011). Additionally, in the United States, self-affirmed White participants reported perceiving more racism, expressed greater belief that White Americans deny racism, and rated the average White person as more racist than did participants who had not been induced to self-affirm (Adams et al., 2006). Indeed, affirmation can enable people to be more open to critical narratives about their in-group’s roles (Sherman, Brookfield, & Ortosky, 2017).

Social-identity threat also stems from out-groups, and a salient example involves in-group members perceiving out-group immigrants as having a different cultural worldview from the majority and, thus, the potential to affect the economy and welfare of the in-group host country (Badea, Iyer, & Aebischer, 2018; Stephan, Diaz-Loving, & Duran, 2000). The hostility toward and prejudice against immigrants stems in part from the external out-group threats (Rothgerber, 1997) they pose to the in-group. See Table 2 for examples of out-group threats.

Our research in France demonstrated that the self-threatening aspect immigrants can present to in-group members may account in part for the prejudice exhibited toward them (Badea, Binning, Verlhic, & Sherman, 2018). Moreover, the expression of prejudice may result partly from having to manage both the symbolic and realistic threats people may believe immigrants pose to their countries and, indirectly, themselves. However, affirming important personal values can secure people’s sense of self-integrity, thus enabling them to respond to perceived threats without the need to bolster the self through prejudice. For example, French undergraduates who wrote about their important personal values perceived less threat, both symbolic and realistic, coming from immigration and consequently displayed reduced prejudice (Badea, Binning, et al., 2018).

### The Effectiveness of Self- Versus Group Affirmation in Reducing Prejudice

Researchers have extended self-affirmation theory by comparing two strategies to affirm self-integrity, namely self-affirmation and group affirmation. The first strategy involves allowing members of the threatened group to bolster values and resources central to the self, whereas the second involves allowing them to affirm values or resources central to a collective identity (Sherman, Kinias, Major, Kim, & Prenovost, 2007). To the extent that immigration threat affects the personal and social self (e.g., participants can imagine their personal and collective economic welfare being harmed by immigrants), self- and group affirmation can be successful at bolstering self-integrity. However, not all variants of affirmation are effective at reducing in-group bias toward and prejudice against immigrants and other minorities.

Whereas self-affirmation has reduced prejudice across a wide range of domains (see Sherman et al., 2017, for review), the pattern of results with group affirmation is less clear: It can attenuate, have no impact on, or even accentuate negative attitudes. For example,

### Table 1. Sample of Prejudice-Reduction Affirmation Studies With Different Types of Responses to In-Group Threats

<table>
<thead>
<tr>
<th>Study</th>
<th>In-group threat</th>
<th>Defensive reaction</th>
<th>Affirmation procedure</th>
<th>Major outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Čehajić-Clancy, Effron, Halperin, Liberman, &amp; Ross (2011, Study 3)</td>
<td>In Bosnia, genocide committed against Muslims</td>
<td>Denial of in-group responsibility</td>
<td>Participants (Serbs) described a personal success (self-affirmation) or a success by a group with which they identified (group affirmation).</td>
<td>Acknowledgment of Serbs’ responsibility for the Srebrenica genocide; feelings of collective guilt (in self-affirmation condition).</td>
</tr>
<tr>
<td>Čehajić-Clancy et al. (2011, Study 1)</td>
<td>In Israel, victimization of Palestinians</td>
<td>Claims that unfortunate events are inevitable in group conflict; citations of serious transgressions committed by Palestinians in other conflicts</td>
<td>Participants (Israeli) ranked a list of values in order of importance to them personally (self-affirmation) or to Israeli society (group affirmation).</td>
<td>Willingness to agree with statements about Palestinians’ victimization and Israelis’ culpability (in self-affirmation condition).</td>
</tr>
<tr>
<td>Gunn &amp; Wilson (2011, Study 2)</td>
<td>In Canada, mistreatment of aboriginal children</td>
<td>Diminishing in-group culpability</td>
<td>Participants (Canadians) ranked a list of values in order of their importance to Canadians (group affirmation).</td>
<td>Collective guilt about Canada’s mistreatment of aboriginal children in residential schools (in group-affirmation condition).</td>
</tr>
</tbody>
</table>

*Each study shown here also included a control condition in which no affirmation was given.*
Canadians who wrote about why a most important Canadian value was important to Canadians (group affirmation) acknowledged greater collective guilt about Canada’s mistreatment of aboriginal children in residential schools, compared with participants in a control condition (Gunn & Wilson, 2011). However, other research (Čehajic-Clancy et al., 2011) found that only self-affirmation was successful in increasing Serbs’ recognition of atrocities committed by Serbs against Bosnian Muslims; group affirmation had no effect. Indeed, it seems that thinking about in-group values and in-group wrongdoings simultaneously causes conflict, and by highlighting group boundaries and divisions, group affirmation may exacerbate defensiveness and biases. In a different study (Ehrlich & Gramzow, 2015), when participants affirmed a value important to their political party, the more they identified with their in-group, the more negatively they evaluated the out-group. Additional evidence suggests that the group-affirmation effect may have been driven by an increase in identity salience. That is, group affirmation increased the accessibility of thoughts related to political party belonging and predicted the group members’ negative evaluations of the opposite party (Ehrlich & Gramzow, 2015).

Accordingly, one plausible explanation for the failure of the group-affirmation procedure to reduce prejudice is that it increases the salience of social identity. Self-affirmation has been shown to have differential effects on openness to identity-threatening information as a function of what identity is made salient in the environment (Cohen et al., 2007), and similarly, group affirmation is likely to fluctuate as a function of what norms are salient within a context. Reflecting on a positive aspect of a social category via group affirmation can increase group members’ identification with their group and, consequently, the motivation to protect its related social identity (Tajfel & Turner, 1979). Also, unlike self-affirmation, group affirmation is often made in the same domain as the threat (e.g., thinking about national in-group values while feeling a threat to that same in-group). Prior research has shown that affirmations in the same domain lead to greater bias (e.g., Sivanathan, Molden, Galinsky, & Ku, 2008) as people focus on defending that domain (Sherman & Hartson, 2011). In contrast, when alternative identities—such as personal identity or a different social identity—are made salient, people may use those resources to confront threats to other aspects of their given collective identity.

### Group Affirmation and Cultural Norms

In some situations, as noted before, group affirmation can foster more conciliatory intergroup attitudes (e.g., Gunn & Wilson, 2011). One important factor that should be taken into account is the normative context when group affirmation occurs. In Canada, for example, group affirmation occurred in a context (college

<table>
<thead>
<tr>
<th>Study</th>
<th>Out-group threat</th>
<th>Defensive reaction</th>
<th>Affirmation procedure</th>
<th>Major outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Badea, Binning, Verlhiac, &amp; Sherman (2018)</td>
<td>In France, perceived terrorist threat coming from immigration</td>
<td>Prejudice against immigrants; support for discriminatory antiterrorism policies</td>
<td>Participants (French) ranked a list of values in order of importance to them personally (self-affirmation) or to them as French citizens (group affirmation).</td>
<td>Perception of lower threat coming from immigrants; lower support for discriminatory policies undertaken by the French government against terrorism (in self-affirmation condition)</td>
</tr>
<tr>
<td>Badea, Tavani, Rubin, &amp; Meyer (2017)</td>
<td>In France, cultural and economic threat coming from immigration</td>
<td>Refusal to welcome Syrian refugees</td>
<td>Participants (French) ranked a list of values in order of importance to them personally (self-affirmation) or to them as French citizens (group affirmation).</td>
<td>Positive intentional behaviors toward refugees (in self-affirmation condition)</td>
</tr>
<tr>
<td>Craig, DeHart, Richeson, &amp; Fiedorowicz (2012)</td>
<td>In the United States, exposing women to reports of sexism by males and society</td>
<td>Pro-White explicit and implicit intergroup bias</td>
<td>Participants (White women) read about the recent success of a popular campus fundraising event organized by their university (a group-affirming event).</td>
<td>Lower intergroup racial pro-White bias after exposure to sexism (in group-affirmation condition)</td>
</tr>
</tbody>
</table>

*Each study shown here also included a control condition in which no affirmation was given.*
Individual differences in value orientation (e.g., individualism vs. collectivism) can influence the effect of self-affirmation on intergroup attitudes. For example, in the aftermath of terrorist attacks in France, people who scored higher in individualism felt more threatened by immigrants yet were more buffered by self-affirmation, and thus they diminished their support for the government’s discriminative policy against immigrants (Badea, Binning, et al., 2018). The tendency to value uniqueness and separateness from other people may leave individualistic people without the psychological buffer that is known to come from being a member of a tight social network (Jetten, Haslam, & Alexander, 2012; Kim, Sherman, & Updegraff, 2016). Therefore, individualistic people may be more psychologically vulnerable to threat and, by extension, more responsive to affirmation (Sherman, Bunyan, Creswell, & Jaremka, 2009).

In other research (Badea, Tavani, Rubin, & Meyer, 2017), French individuals indicated greater willingness to welcome immigrants after self-affirming values congruent with their political orientation (e.g., politically left-oriented participants wrote about left-wing values such as equality). In other words, self-affirmations that were congruent with individuals’ values orientation were more effective in reducing in-group bias and prejudice compared with incongruent self-affirmations. Such studies reaffirm this general theoretical principle: Self-affirmations that tap into the values that are central to the individual are likely to be the most affirming. We also note that group affirmations can sometimes be sensitive to individual differences (e.g., in-group identification; Sherman et al., 2007), but more research is needed to identify the weighting of normative and individual factors in the efficacy of group-affirmation manipulations.

**Potential for Longer-Term Prejudice Reduction**

One exciting implication of the self-affirmation approach in the domain of prejudice reduction is that self-affirmation shows the potential malleability of prejudice in situations of intergroup conflict (Sherman et al., 2017). Applications of self-affirmation theory in prejudice reduction may be developed, as affirmations have demonstrated long-term beneficial effects in educational settings when they foster adaptive recursive processes among people experiencing identity threat (Cohen & Sherman, 2014). Evidence suggests that, in itself, the very process of affirmation leading to academic improvement can become self-affirming (Cohen, Garcia, Purdie-Vaughns, Apfel, & Brzustoski, 2009).

For affirmation to lead to long-term prejudice reduction, the intervention must lead to some change in the individual that could be reinforced by the cultural context in which the person resides; otherwise, the effects are likely to be short-lived (Sherman et al., 2017). Thus, if affirmation can foster a beneficial interaction between people who typically exhibit prejudice and stereotyping toward each other, then the beneficial interaction itself could become its own source of affirmation, perhaps leading spirally to more impactful effects.

Charlotte Brontë (1847/1864) wrote that “prejudices, it is well known, are most difficult to eradicate from the heart whose soil has never been loosened . . . they grow there, firm as weeds among stones” (p. 361). Prejudices are difficult to eradicate because they are multiply determined. Consequently, an individual-level solution that focuses solely on the self is, of course, not
Self-Affirmation and Prejudice Reduction

a sufficient means to reduce prejudice. Rather, addressing the structural antecedents of prejudice, such as inequality, fear, division, and competition for resources, is imperative (Bar-Tal & Halperin, 2011; Mahfud, Badea, Verkuyten, & Reynolds, 2018; see also Richeson, 2018, for a discussion in the context of racism). As demonstrated, the motivation to maintain feelings of self-worth and self-integrity can lead to prejudice, but the same motivation may also inform approaches that could lead to reducing prejudice. Clearly, there is a psychological element of prejudice that stems from the perception of self-threat. Perhaps this could be addressed through affirming activities that may, in turn, lead to “a loosening of the soil” in which prejudice takes root.

**Recommended Reading**


**Action Editor**

Randall W. Engle served as action editor for this article.

**ORCID iD**

Constantina Badea [https://orcid.org/0000-0001-7459-4698](https://orcid.org/0000-0001-7459-4698)

**Acknowledgments**

The authors thank Kevin Binning, Phil Ehret, Kimin Eom, Lauren Ortosky, Rammy Salem, Michelle Shteyn, Payton Small, and Hilary Watts for their comments on this manuscript.

**Declaration of Conflicting Interests**

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

**References**


