

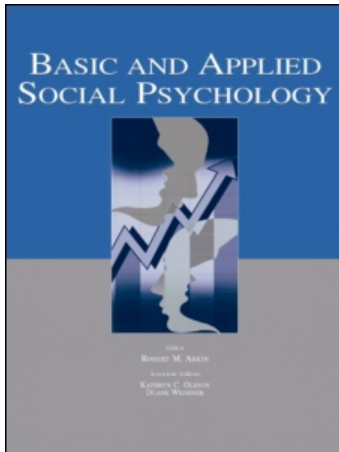
This article was downloaded by: [Conway, III, Lucian Gideon]

On: 19 November 2009

Access details: Access Details: [subscription number 916999026]

Publisher Psychology Press

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



## Basic and Applied Social Psychology

Publication details, including instructions for authors and subscription information:

<http://www.informaworld.com/smpp/title~content=t775648090>

### When Self-Censorship Norms Backfire: The Manufacturing of Positive Communication and Its Ironic Consequences for the Perceptions of Groups

Lucian Gideon Conway III <sup>a</sup>; Amanda Salcido <sup>a</sup>; Laura Janelle Gornick <sup>a</sup>; Kate Ashley Bongard <sup>a</sup>; Meghan A. Moran <sup>a</sup>; Chelsea Burfiend <sup>a</sup>

<sup>a</sup> The University of Montana,

Online publication date: 18 November 2009

**To cite this Article** Conway III, Lucian Gideon, Salcido, Amanda, Gornick, Laura Janelle, Bongard, Kate Ashley, Moran, Meghan A. and Burfiend, Chelsea(2009) 'When Self-Censorship Norms Backfire: The Manufacturing of Positive Communication and Its Ironic Consequences for the Perceptions of Groups', Basic and Applied Social Psychology, 31: 4, 335 – 347

**To link to this Article:** DOI: 10.1080/01973530903317169

**URL:** <http://dx.doi.org/10.1080/01973530903317169>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

# When Self-Censorship Norms Backfire: The Manufacturing of Positive Communication and Its Ironic Consequences for the Perceptions of Groups

Lucian Gideon Conway, III, Amanda Salcido, Laura Janelle Gornick, Kate Ashley Bongard, Meghan A. Moran, and Chelsea Burfiend

*The University of Montana*

Do norms compelling self-censorship of negative communication work? An attributional analysis suggests that awareness of self-censorship norms causes people to be suspicious of other people's positive communications about groups, thus causing the norms to backfire. Three studies tested this *informational contamination* hypothesis. Participants read stories in which they imagined that some friends' conversations painted a particular fraternity in a good light. Results from all three studies revealed that when participants were exposed to a cue encouraging self-censorship—the presence of a member of the talked-about fraternity—this self-censorship norm backfired, instead leading them to talk disparagingly about the fraternity in a different context. Mediation analyses implicate an informational contamination explanation for the backfiring effect more so than a reactance-based explanation.

People often don't say what they are really thinking. A woman might tell a date that she really likes his new leather jacket even though she privately hates it. Later that evening, they might both tell the waiter that they liked the food even though it was cold and poorly prepared.

People often don't say what they are really thinking about leather jackets and undercooked food because they want to conceal a negative opinion that might cause themselves or others discomfort. Indeed, much research illustrates that people are often motivated to present themselves in a light that does not fully represent their real opinions and that these self-presentational concerns have multiple consequences (e.g., Arkin, Gabrenya, Appelman, & Cochran, 1979; Baumgardner, 1991; Conway et al., 2008; Feldman, Forrest, & Happ, 2002; Kolditz & Arkin, 1982; Schaller & Conway, 1999; Tyler & Feldman, 2005).

## SELF-CENSORSHIP OF GROUP-RELATED COMMUNICATION: "POLITICAL CORRECTNESS" NORMS

Political correctness is a term applied to language, ideas, policies, or behavior seen as seeking to minimize offense to gender, racial, cultural, disabled, aged or other identity groups. (Wikipedia, 2008)

These self-presentational concerns are especially strong when the concealed opinion might cast a negative shadow over an ethnic, gender, or political group. Indeed, as implied in the quote just presented, an extremely powerful set of norms governing disparaging references to groups has emerged in modern North American society. This set of norms, often called political correctness (or PC) norms, explicitly attempts to remove negative group-relevant language. As a result, in situations where the norms are in evidence, they create a particularly strong motive for self-censorship. Even when people may really believe that a female manager is not very talented, they are especially unlikely to give her a negative review—for fear that it will violate the norm.

Although much derided (indeed, the very term *political correctness* has come to have a negative connotation,

---

Correspondence should be sent to Lucian Gideon Conway, III, Department of Psychology, The University of Montana, Missoula, MT 59812. E-mail: luke.conway@umontana.edu

often referring particularly to cases where it seems to have gone overboard), few academics would disagree that the practical goal of the PC movement is well-aimed. The movement explicitly seeks to reduce unwarranted negative stereotypes of often disadvantaged groups. We applaud the goal. But it is a further question to ask: Will the PC movement achieve that goal?

The purpose of this article is to discuss the applied consequences of such self-censorship norms on the ultimate contents of group stereotypes. We review reasons why these norms might work, and some additional reasons why they might not work so well. Then we report a critical test of the two ideas that illuminates a context in which making self-censorship norms salient actually backfires, resulting in more disparaging remarks about the group.

### WHY SELF-CENSORSHIP NORMS MIGHT SUCCEED: A COMMUNICABILITY PERSPECTIVE

Some theory and research suggests that self-censorship norms may succeed in reducing unwarranted negative stereotypes. One perspective, for example, suggests that stereotypes are like viruses: They spread because they are highly communicable, and they die out when they cannot be communicated from host to host (e.g., Conway & Schaller, 2007). Thus, pressures that make negative stereotypes harder to talk about (as self-censorship pressures do) may necessarily reduce their chance of survival. For example, Schaller and Conway (1999) demonstrated that when subtle pressure exists to talk in a positive light about groups, persons not only comply with the pressure and talk more positively, but also form more positive stereotypes as a result of these positive interactions. The impact of communication on stereotypes isn't just a laboratory phenomenon, either: Other work suggests that the development of stereotypes of real ethnic groups over a 60-year period was shaped by how likely persons were to talk about various stereotypic traits (Schaller, Conway, & Tanchuk, 2002). This evidence cumulatively suggests that if norms exist encouraging people to talk positively about groups, people will indeed talk—and eventually believe—more positively about groups. Thus, making such self-censorship norms salient may inevitably lead to an increase in the overall positivity of group-relevant communication and beliefs.

### WHY SELF-CENSORSHIP NORMS MIGHT BACKFIRE: AN ATTRIBUTIONAL PERSPECTIVE

Although it is almost certainly true that invoking a self-censorship norm will make people more likely to

communicate positively in the immediate circumstance, it may (sometimes less obviously, but no less importantly) have other, more insidious consequences as well. For people do more than talk during conversations; they also hear others talk as well. What are the consequences of self-censorship norm salience for the listeners?

An attributional perspective suggests an answer (Conway & Schaller, 2005). When perceivers believe that others' expressed opinions are constrained by the operation of a social norm or other external constraints, they are less likely to attribute those expressions to those others' actual opinions (Conway & Schaller, 2005; Fein, 1996; Fein, Hilton, & Miller, 1990; Harkins & Petty, 1987; Pryor, Rholes, Ruble, & Kriss, 1984; Wilder, 1977; see also Gurevitch, 1985). For example, when observers believe that others are expressing an opinion because they were commanded to do so by an authority figure, observers don't believe their opinion is sincere—and as a result are less likely to obey the authority's command themselves (for evidence from business scenarios, see Conway & Schaller, 2005). The informational value of the expressed opinions we hear from others becomes contaminated by the authority figure's command.

Similarly, a norm imploring us to censor our own words about groups may cause us to indeed censor our words, but it also may cause us to question (where otherwise we would not have questioned) the legitimacy of other people's expressed positive opinions about those same groups. It contaminates any subsequently positive opinions we may hear expressed, and it does so even if those opinions of others may have been perfectly genuine. For example, a set of friends might be talking positively about a particular fraternity because they actually all think the fraternity is great. All else being equal, someone listening to that conversation would—due to the power of a consensually expressed opinion—be persuaded to think the fraternity is more positive than they had previously believed, and as a result the observer would be more likely to express a positive opinion of the fraternity themselves if later given the chance. But if the observer becomes aware that a member of the fraternity was present during the original conversation—suggesting that the social norm for self-censorship about groups was particularly salient—they may believe instead that no one really believes positively about the fraternity after all (“They just felt like they had to say that; they didn't mean it”). Given the opportunity later in a different context, this observer might talk more disparagingly about the group than they otherwise would have. As a result, the salience of a self-censorship norm can actually lead to more disparaging talk about the group.

In a sense, a self-censorship norm, like the command of an authority figure (Conway & Schaller, 2005), is a double-edged sword: It artificially creates a positive

consensus among people who feel immediately compelled to comply with the norm, but it also undermines the psychological power of that positive consensus in the minds of others who perceive their opinions. As a means of reducing negative stereotypes, then, self-censorship norms may backfire by contaminating the informational value of positive communications, and thus making those communications less compelling.

### Emotional Versus Informational Reasons for Deviance: Reactance Theory and Attribution

One of the implications of this perspective is that pressures to conform to group norms can backfire because they contaminate the subsequently communicated information. Of course, such conformity pressures can backfire for a number of reasons that have nothing to do with informational contamination (Knowles & Linn, 2004). Perhaps the most prominent example is found in psychological reactance theory (e.g., Brehm, 1966; Crawford, McConnell, Lewis, & Sherman, 2002; Fiegen & Brehm, 2004). Real or perceived pressure from others to engage in a particular behavior often causes us to feel a loss of our personal freedom to choose. Freedom of choice is a valued psychological commodity, and so reactance theory posits that we often will deviate from others' expectations—both in belief and in action—to reassert our right to choose. Thus, like the attributional model, reactance theory also predicts that under some circumstances a norm compelling us to engage in a shared behavior will inspire deviance.

Although predicting some of the same outcomes, the reactance and attributional approaches provide quite different accounts of the origins of such normative deviance. The reactance approach is a more purely affect-driven account of deviance (Knowles & Linn, 2004). The attribution approach, in contrast, is more purely cognitive: It suggests that persons deviate because of factors that contaminate the informational value on which the influence of others is built. No feeling of reactance—indeed, no feeling at all—is necessary for this contamination to occur (Conway & Schaller, 2005).

These two possible processes are psychologically distinct, and so are not necessarily “competing” explanations. The present studies are designed primarily to test hypotheses derived from the attributional approach. Supportive results, of course, cannot imply that norms do not ever backfire for emotion-based responses such as reactance. Affective processes are clearly important in understanding the origins of deviant behavior. But even important psychological processes are not all-encompassing; the operation of attributional processes may explain deviance phenomena that reactance

(and other affective processes) cannot (see Conway & Schaller, 2005). Indeed, to test the unique predictive validity of the two approaches in our own work, in Study 3 we include measurements of both informational contamination and reactance. As we see, although both approaches have mediation power, the informational contamination processes account for more unique variance.

### Possible Moderating Variables

Although self-censorship norms may backfire, the attributional model does not necessarily suggest that this will occur in all circumstances. We test two potential moderating variables in the present work. Study 1 tested the moderating power of cognitive load, whereas Study 3 tested the moderating power of the immediate presence of the self-censorship cue. As we see, although (consistent with the informational contamination hypothesis) all three studies showed a backfiring effect of the self-censorship cue, neither moderator approached statistical significance. Of interest, mediation analyses suggest that informational contamination in some sense was so powerful that it simply overrode these potential moderators in these contexts.

## RATIONALE AND OVERVIEW OF THE PRESENT STUDIES

In Studies 1 to 3, we aim first and foremost to provide evidence that, in spite of the obvious pressures that might lead self-censorship norms to reduce negativity, attributional processes can cause self-censorship norms to lead to *more* expression of negative opinions of groups. We thus set up a circumstance in all three studies reported here where the attributional model would predict exactly this kind of backfiring effect. We applied a scenario methodology similar to that used in previous attribution research (Conway & Schaller, 2005; Fosterling, 1989; Hewstone & Jaspars, 1987; Kelley, 1967; McArthur, 1972), previous stereotyping research (e.g., Alexander, Brewer, & Herrmann, 1999; Eagly & Wood, 1982), and prior specific research on our targeted “stereotyped” group, fraternities (Drout & Corsoro, 2003; Indick, 2003; Miles & Naumann, 2003). Participants in our studies read a short story about a group of university friends who are discussing a campus fraternity in a positive manner. All three studies contained a manipulation (to which the experimenter was blind) pertaining to the degree that a self-censorship norm was salient: Whether a member of the fraternity was present or absent during this discussion. We reasoned that the presence of the talked-about fraternity member would heighten participants' recognition that it was socially inappropriate to

talk badly about the fraternity in that context. Later in the story, the participants' "character" was talking to another friend, and participants were asked to complete (open-ended and rating scale) measures indicating what they would communicate to this other friend about the fraternity if they themselves were in this situation. Participants in all studies also completed measures relevant to their attributions about why the popularly expressed opinion that the "fraternity is great" existed, and (in Study 3) about the degree to which they felt reactance.

Based on the attributional model, we expected that (under the right circumstances) the presence of the self-censorship cue would backfire, ultimately making communication more negative. We also expected that this backfiring effect would be accounted for in part by the attributions that participants made concerning why others had communicated positively about the group in the first place. Thus, after we present all three studies here, we present some pooled mediation analyses concerning the degree to which our attribution measures (as well as reactance measures) accounted for the key backfiring effect.

## STUDY 1

### Overview

Study 1 participants read a short scenario in which they imagined themselves listening to a group of friends talking. This group of friends talks in a positive manner about a campus fraternity (the actually fictitious "Sigma Sigma Sigma" fraternity). Before they do so, a person is introduced to the group as either a member of the talked-about fraternity (Self-Censorship Cue Present) or an already-known member of the circle of friends (Cue Absent condition). Participants are then asked what they would communicate about the group to a different friend in another context. Further, previous research (Conway & Schaller, 2005, Study 5) suggests that other ironic effects associated with attributions about consensus require the availability of cognitive resources; when such resources were lessened through a distraction task, participants did not show the key effect. Although the relationship between effortful thinking and group perceptions can be complicated (see, e.g., Gilbert & Hixon, 1991; Wegener, Clark, & Petty, 2006), the present study manipulated cognitive load to see if it similarly moderated the backfiring effect.

As we report later, Study 1 did provide strong support for the basic backfiring effect but showed no moderating impact of cognitive load. We further address these issues in the discussion.

## Method

### Participants

Two hundred fifteen introductory psychology students at Indiana State University participated for course credit in several mass testing sessions. Sixteen participants who did not make an effort to memorize the eight-digit number were dropped, leaving 199 for final analyses (inclusion of these additional participants does not change any of the results reported later, either descriptively or inferentially).

### Stories

Participants were randomly assigned to read one of two stories (either the self-censorship cue present story or the cue absent story). Regardless of condition, participants were asked to imagine themselves in a scenario (described later) and to think about what they would say or do if they were in this scenario.

**Introduction.** All participants were asked to imagine overhearing a conversation between a group of friends in the University Center. This introduction emphasized how diverse the set of friends appeared to be and how unusual it is for them all to agree on any one topic. The primary goal for this segment was to ensure that the resulting expressed opinion cannot be easily attributed to some commonly shared background or personality trait.

**Self-censorship cue.** After the story introduction, the group in the scenario was always introduced to "Mary's" friend "Steve." The self-censorship manipulation pertained to whether Steve was described as a member of the fraternity itself or as a regular member of the circle of friends. In particular, self-censorship cue participants read the following:

"Oh, hi, Mary. Glad you could come," said Jim. "Who's your friend?"

Mary introduced the new guy. "Hey, everyone," Mary said, "I'd like you to meet my new friend Steve. I don't think any of you have met him before. Steve's a member of the Sigma Sigma Sigma Fraternity."

It was obvious that, in fact, none of them had ever met Steve before. After introductions had been made, Jim picked up the conversation he had started.

However, no cue participants read the following:

"Oh, hi, Mary. Glad you could come," said Jim. "And hi Steve."

It was obvious that both Mary and Steve were well-known to everyone in the group. After they all said hello, Jim picked up the conversation he had started.

**Cognitive load manipulation.** In addition, cognitive load was manipulated in a manner similar to previous research (e.g., Conway & Schaller, 2005). Half of the participants received written directions instructing them to try to remember a given eight-digit number as they read the story. The other half of participants did not receive these directions.

### *Dependent and Attribution Measures*

**Free response measures.** Immediately after reading the story, participants were asked to complete two free response questions. The first question, used to assess disparaging communication intent, was, "If you had experienced the preceding scenario, what do you think you would say to your friend about the Sigma Sigma Sigma Fraternity in this situation? Write out a response to your friend." These responses were coded, blind to condition, along a bipolar dimension using a 1-to-9 scale. These ratings answered the question, "How positive/intelligent does this response suggest the Sigma Sigma Sigma fraternity is?" where 1 (*very negative/dumb*) to 9 (*very positive/intelligent*), and 5 (*an equal combination of the two*).

The second question, "Why do you think Bill's group of friends all expressed the same opinion toward the Sigma Sigma Sigma fraternity?" assessed attributions about consensus. These were coded, blind to condition, on a bipolar dimension using a 1-to-9 scale. These ratings answered the question, "How much is the shared opinion that the Sigma Sigma Sigma fraternity is intelligent attributed to actual reality (or the real experience of group members) versus social influence (e.g., presence of the fraternity member, group members' influence)?" from 1 (*social influence*) to 9 (*reality*), with 5 (*an equal combination of the two*).

For each of these two items, a second coder scored a small subset of the responses; the two coders' ratings were highly correlated for both communication intent ( $r = .93$ ) and attributions ( $r = .98$ ).

**Nine-point rating scales.** In addition to these free response questions, other questions (9-point rating scales) assessing both communication intentions and attributions were completed by participants. Some of these questions pertained to participants' likelihood of expressing specific views of the Sigma Sigma Sigma fraternity to their friend: For example, "How likely is it that you would tell your friend that the Sigma Sigma Sigma Fraternity members are really intelligent?" Similar questions were completed relevant to communicating "not very intelligent," "aggressive," "positive," and "negative" impressions of the fraternity.

An additional set of items that pertained to attributions about consensus also was completed by participants: For

example, "Based on the story, to what degree do you think the fact that the circle of people in the story all agreed when they expressed their opinion of the Sigma Sigma Sigma Fraternity is because that's what they really believed?" Similar questions asked participants to what degree they attributed the circle of friends' agreement to the real opinions of group members, concern about outwardly saying negative things, and conformity processes, respectively.

### *Construction of Measures*

**Disparaging communication intent.** There was a general tendency for the measures of communication intent to be correlated with each other. To create a general summary measure of overall communication intent, all seven measures of communication intent (one free response intelligence/positivity item, three 9-point intelligence items, two 9-point general positivity items, and one 9-point aggressive item), were converted to  $z$  scores, reversed scored (when necessary) so that higher scores always represented more negative communication intent, and averaged (for these scores, intelligence was always treated as a positive trait and aggressive as a negative trait; see Rothbart & Park, 1986; Schaller & Conway, 1999). This *disparaging discussion intent* score (standardized  $\alpha = .81$ ) represented the degree that participants reported that they would communicate a negative impression of the Sigma Sigma Sigma fraternity to their friend in the story and served as the primary dependent measure in the analyses.

**Attributions.** Last, an attribution  $z$  score composite was created from the five measures that were relevant to attributions about the communications amongst the circle of friends. When necessary, items were reversed scored so that higher numbers always meant more suspicion of external influences. This composite (standardized  $\alpha = .71$ ) represented the degree that persons attributed the agreement of the friends to external influences that do not represent persons' actual experience with group members. This *attribution* composite served as the primary mediator in the mediation analyses (see pooled analyses after Study 3).

## Results

Primary analyses were conducted within a 2 (self-censorship cue: yes or no)  $\times$  2 (cognitive load: high or low) design.

### *Attributions*

The Self-Censorship manipulation had its intended effect on attributions. Consistent with expectations,

participants were more likely to attribute the positive statements about the Sigma Sigma Sigma fraternity to something other than the communicators' real beliefs when a fraternity member was present ( $M = .21$ ) than when he was absent ( $M = -.21$ ),  $F(1, 195) = 8.90$ ,  $p < .001$ . Thus, the presence of the fraternity member made people more suspicious of the positive statements others made about his fraternity.

A main effect emerged for cognitive load as well: Participants were more suspicious of the positive statements of others in no load than in load conditions ( $p = .022$ ); thus, load made participants less likely to question the sincerity of others' expressed opinions. However, no Load  $\times$  Cue interaction emerged ( $p > .73$ ). Please see Table 1.

### Disparaging Discussion

Analyses of the primary dependent measure yielded results consistent with the informational contamination hypothesis. In particular, a main effect emerged for self-censorship cue: Participants were more likely to later talk disparagingly about the fraternity when a fraternity member was present at the initial discussion ( $M = .12$ ) than when he was absent ( $M = -.09$ ),  $F(1, 195) = 4.53$ ,  $p = .035$ , effect size (Pearson's  $r$ ) = .14. No main effect of cognitive load, or a Load  $\times$  Cue interaction, emerged ( $ps > .41$ ). See Table 1.

### Discussion

Study 1 provided initial evidence that a self-censorship norm can backfire. Instead of producing (in line with its intent) more positive discussions about groups, a salient self-censorship norm ironically produced more disparaging communication.

It was anticipated, based on previous research (Conway & Schaller, 2005, Study 5), that the backfiring effect would be weakened when participants were under cognitive load. However, whether or not participants were under cognitive load, the self-censorship cue

increased the amount of disparaging discussion. Why might this be? One possibility for why participants' final judgments were more impacted by cognitive load in previous work is that, in that work, the ultimate judgment was a more difficult cognitive task. In previous work, participants had to report their likelihood of a final decision between two different systems, whereas in the present study participants had to report only what they would say about one talked-about group. Not only does the first scenario involve more categories to consider (two vs. one), it also may be cognitively "farther away" from the attributional judgment serving as the basis for the backfiring effect. As a result, cognitive load may have interfered with these judgments in prior work, but because the judgment required in the current work is cognitively "easier," participants may have still been able to perform it identically in load (vs. no load) conditions.

At any rate, whatever the reason for cognitive load's lack of moderation of the key effect, it serves as a potential qualifier on previous research; the interactive effect of cognitive load and attributional cues relevant to consensus may not be particularly robust. Of course, this qualifier does not in any way undermine the consistent ironic effects across vastly different domains that considering attributions about consensus has produced.

## STUDY 2

One possible interpretation of the results of Study 1 presented is that it is the presence of a stranger, and not the presence of a fraternity member specifically, that causes the backfiring effect. Although this would not be inconsistent with the current informational contamination approach (anything that possibly heightens one's sense of self-censorship norms ought to invoke a backfiring effect potential), it is worth understanding the exact locus of the effect. To that end, Study 2 adds an additional control condition where the person introduced to the group is a stranger but not a member of the talked-about fraternity.

### Overview

Study 2 was identical to Study 1 in most respects. Three changes were evident. First, instead of experiencing a fully first-person scenario, participants read about a character ("Bill") who is placed in a scenario virtually identical to the scenario from Study 1. Participants were instructed to "place themselves in Bill's shoes" and subsequently answered all questions as if they were in the scenario in Bill's place. Functionally, this is the same as placing themselves in the scenario (as in Study 1)—and, in fact, results demonstrated an identical pattern.

TABLE 1

Study 1: Effects of Self-Censorship Cue and Cognitive Load on Attributions and Disparaging Communication Relevant to the Fraternity

	No Load		Load	
	No Cue	Cue	No Cue	Cue
Attribution	-.13	.33	-.31	.08
Disparaging Communication	-.04	.14	-.15	.09

Note.  $N = 199$ . Higher scores on attribution measure = more suspicion of informational contamination; higher scores on disparaging communication = more intent to communicate negatively about the fraternity.

Second, Study 2 did not include a cognitive load manipulation. Third, Study 2 added an additional control condition: In addition to the cue condition where a Sigma Sigma Sigma fraternity member is present at the initial conversation and the no cue/friend condition where a friend is present, the present study also included a no cue/stranger condition where a person who is neither a member of the fraternity nor a friend of the group is present.

## Method

### *Participants*

Fifty-one participants from the University of British Columbia participated in the study for course credit.

### *Stories*

Participants were randomly assigned to read one of three stories designed to match each of the three conditions. These stories were virtually identical to Study 1, including an overview of the diversity of the group, the information about the Sigma Sigma Sigma fraternity, and the subsequent conversation.

The cue and the no cue/friend conditions were virtually identical to those from Study 1. In the no cue/stranger condition, participants read that the new person ("Steve") introduced to the group and present during the subsequent conversation is a stranger who is a member of the "University Golf Club."

This no cue/stranger condition allows finer distinctions to be made as to the exact locus of any emergent effect of the self-censorship cue. In particular, it allows us to disentangle whether any emergent difference between the no cue/friend and cue conditions is due to the presence of a stranger in general, versus the specific presence of a stranger in the fraternity under discussion in the story.

### *Dependent and Attribution Measures*

All dependent and attribution measures were identical to those in Study 1, with the exceptions that (a) they were adapted to the perspective-taking of the story (e.g., "if you were in Bill's shoes, what would you have said to your friend?") and (b) one (out of seven) of the 9-point rating scales for disparaging discussion was not used. As before, the two free response tasks were coded for communication negativity and attribution content, and interrater reliability was satisfactory (reliabilities = .80 and .87). Composite measures were also constructed in a manner identical to the previous studies: A measure representing participants' attributions about consensus (standardized  $\alpha = .87$ ) and a measure representing participants' disparaging discussion intent with respect to the fraternity (standardized  $\alpha = .74$ ).

## Results

### *Analytic Strategy*

Two parallel sets of analyses are presented below. First, one-way analyses of variance (ANOVAs) that reveal the omnibus likelihood that differences among any of the three conditions would have emerged due to sampling error are presented. Second, to directly replicate the effect from Study 1, a more focused planned comparison between the cue and no cue/friend conditions is also presented. Equally as importantly, the descriptive pattern of results with respect to the no cue/stranger condition will help us understand the precise locus of the self-censorship cue's effect.

### *Attributions*

The self-censorship cue manipulation had its intended effect on attributional processes. A one-way ANOVA revealed that the mean differences emerging between the three conditions were not likely due to sampling error,  $F(2, 48) = 7.90, p < .001$ . A planned comparison between the cue and no cue/friend conditions revealed that, as expected, participants in cue conditions reported more attributional contamination ( $M = .46$ ) than persons in the no cue/friend ( $M = -.08$ ) condition,  $t(31) = 2.11, p = .022$ , one-tailed. Consistent with the idea that the specific presence of the fraternity member is necessary to create attributional suspicion, the no cue/stranger condition did not show high scores for attributional contamination (indeed, it showed the lowest score of any condition,  $M = -.48$ ).

### *Disparaging Discussion*

The self-censorship cue manipulation had its intended effect on communication intent. Although the overall ANOVA showed a weak effect of the cue manipulation (one-way ANOVA),  $F(2, 48) = 1.27, p = .292$ , the planned comparison between the Cue ( $M = .21$ ) and no cue/friend ( $M = -.13$ ) conditions revealed a relatively small likelihood that the difference between these two conditions resulted solely from sampling error,  $t(31) = 1.63, p = .057$ , one-tailed, with an effect size even larger than in Study 1 (Pearson's  $r = .28$ ). As with the attribution measure, the no cue/stranger ( $M = -.07$ ) condition showed a pattern of results closer to the no cue/friend condition than to the cue condition.

## Discussion

Study 2 again offered support for the main conceptual hypothesis: As in Study 1, participants who heard a positive consensual view presented of the fraternity were more likely to subsequently disparage the group when



an implicit norm to censor negative communication had existed in the original context. In addition, it appears that the self-censorship cue's backfiring effect requires, not merely that a stranger is present for the discussion, but that a stranger who is a member of the stereotyped group is present. Thus, it is not just the presence of any stranger, but rather the specific presence of a member of the group, that activates the self-censorship norm.

### STUDY 3

#### Overview

Studies 1 and 2 provided evidence that, under the right circumstances, activating a norm designed to make persons censor their negative communications about a group can actually backfire, producing more disparaging communication instead. Study 3 aimed to build on this conclusion by (a) replicating the key backfiring effect, (b) incorporating a reactance measure to determine the degree that reactance mediates this effect (see pooled analyses after Study 3 results), and (c) including a manipulation of the immediate presence of the fraternity member when participants make their communication ratings. Intuitively, one would expect the backfiring effect to be attenuated when persons still feel they are directly in the presence of the self-censorship cue (thus feeling the pressure to abide by it at that moment), in the same way that an authority figure's presence attenuates effects based on informational contamination.

#### Method

##### *Participants*

Three hundred introductory psychology students at the University of Montana participated for course credit in several mass testing sessions.

##### *Stories and Self-Censorship Cue Manipulation*

Participants were randomly assigned to read one of four stories designed to match a 2 (self-censorship cue: present or absent)  $\times$  2 (immediate presence of cue: yes or no) design. These stories were virtually identical to those used in Study 1. In addition, the self-censorship cue manipulation was identical to that from Study 1, with participants in cue present conditions hearing a conversation that took place in the presence of a member of the talked-about fraternity, whereas cue absent conditions heard a conversation that occurred in the presence of a mutual friend.

##### *The Immediate Presence Manipulation*

At the end of the story, participants were asked to communicate their own impressions of the fraternity

to another person (described as their own friend in the scenario). We manipulated the context within which this communication occurred. In the immediate presence condition, participants read the following:

Just then, as you were listening, imagine that you were startled when another person who had just joined the group called out your name. It was a friend of yours that you knew pretty well. "Hey, come over here!" said your friend. As you walk up, your friend – in the presence of everyone in the group you had been "eavesdropping" on, all of whom were now listening only to you—asks you "What do you think of the Sigma Sigma Sigma fraternity?"

In the no immediate presence condition, participants read the following:

Later that week, imagine you were talking to a friend of yours that you knew pretty well. You and your friend began to talk about various organizations on campus. At one point your friend asks you, "What do you think of the Sigma Sigma Sigma fraternity?"

##### *Dependent, Attribution, and Reactance Measures*

All dependent and attribution measures were identical to those in Studies 1 and 2. As before, the two free response tasks were coded for negativity and attribution content, and interrater reliability was satisfactory (reliabilities = .79 and .96). Composite measures were also constructed in a manner identical to the previous studies: A measure representing participants' attributions about consensus (standardized  $\alpha = .83$ ) and a measure representing participants' disparaging discussion intent with respect to the fraternity (standardized  $\alpha = .70$ ).

Further, in Study 3, after completing the other measures, participants also completed three questions pertinent to reactance adapted directly from prior research (Conway & Schaller, 2005). For example, "To what degree did some part of the scenario make you feel as if someone was trying to take away your freedom to do (or believe) exactly as you wished?" Participants were asked similar questions about how much they felt some part of the scenario made them want to do the opposite of implied social pressure to show people that they could not be told what to do, and to what degree the scenario made them upset that someone else tried to impose their will on the participant. These questions were averaged to form a reactance composite measure (standardized  $\alpha = .75$ ).

#### Results

Primary analyses were conducted within a 2 (self-censorship cue: yes or no)  $\times$  2 (immediate presence: yes or no) design.

*Attributions*

The self-censorship manipulation again had its intended effect on attributions. Consistent with expectations, participants were more likely to attribute the positive statements about the Sigma Sigma Sigma fraternity to something other than the communicators' real beliefs when a fraternity member was present ( $M = .28$ ) than when he was absent ( $M = -.28$ ),  $F(1, 296) = 43.68$ ,  $p < .001$ . Thus, the presence of the fraternity member made people more suspicious of the positive statements others made about his fraternity.

A marginal main effect emerged for immediate presence as well: Participants were more suspicious of the positive statements of others in immediate presence ( $M = .07$ ) than in no immediate presence ( $M = -.08$ ) conditions ( $p = .072$ ); thus, being forced to speak in the immediate presence of others made participants more likely to question the sincerity of others' expressed opinions. However, no Presence  $\times$  Cue interaction emerged ( $p > .45$ ). Please see Table 2.

*Disparaging Discussion*

Analyses of the primary dependent measure yielded results consistent with Studies 1 and 2. In particular, a main effect emerged for self-censorship cue: Participants were more likely to talk disparagingly about the fraternity when the cue was present at the initial discussion ( $M = .09$ ) than when it was absent ( $M = -.09$ ),  $F(1, 296) = 6.69$ ,  $p = .010$ , effect size (Pearson's  $r$ ) = .15. No main effect of immediate presence, nor a Presence  $\times$  Cue interaction, emerged ( $ps > .16$ ). See Table 2.

**Discussion**

Results were consistent with Studies 1 and 2, again demonstrating the key backfiring effect: A cue that ostensibly

should produce less disparaging communication instead produced more such communication. Of interest, the immediate presence of the cue did not significantly moderate the effect. Why might this be? The most straightforward explanation involves the relative subtlety of the self-censorship cue. When a fraternity member is present, participants may be aware of the cue enough to make attributional corrections about what others are saying but not concerned enough about it to modify their own statements overmuch (for a similar discussion, see Conway & Schaller, 2005).<sup>1</sup>

To this point, the body of work presented here is almost entirely consistent with the fact that self-censorship cues can actually backfire, and that they do so because they cause a kind of informational contamination. However, an important question relevant to this proposed causal path remains unanswered: Is there direct evidence of attributional mediation of the backfiring effect? It is to these tests that we now turn.

POOLED MEDIATION ANALYSES  
(STUDIES 1–3)

We performed pooled mediation analyses on the data from Studies 1 to 3 to see if attributions about consensus mediated the effect of the self-censorship cue on disparaging communication. In Studies 1 and 3, we collapsed across the cognitive load and immediate presence conditions, as they did not moderate the effect in those studies. Also, in Study 2 we collapsed the two no cue conditions into one dummy variable, because both no cue conditions showed similar effects.

These pooled analyses were consistent with hypotheses. Overall, the zero-order correlation between the self-censorship cue manipulation (coded as 0 = cue absent, 1 = cue present) and the amount of disparaging communication about the Sigma Sigma Sigma fraternity was significant  $r = .15$ ,  $p < .001$ , one-tailed. However, when controlling for attributions about consensus, this relationship was substantially reduced ( $r = .06$ ,  $p = .078$ , one-tailed).

TABLE 2

Study 3: Effects of Self-Censorship Cue and Immediate Presence of Cue on Attributions, Reactance, and Disparaging Communication Relevant to the Fraternity

	<i>No Immediate Presence</i>		<i>Immediate Presence</i>	
	<i>No Cue</i>	<i>Cue</i>	<i>No Cue</i>	<i>Cue</i>
Attribution	-.39	.23	-.17	.32
Reactance	-.42	.10	-.06	.36
Disparaging Communication	-.16	.06	-.02	.11

*Note.*  $N = 300$ . Higher scores on attribution measure = more suspicion of informational contamination; higher scores on disparaging communication = more intent to communicate negatively about the fraternity.

<sup>1</sup>An additional study using a more powerful self-censorship cue supported this conclusion (Conway & Gornick, 2009). Instead of manipulating self-censorship norms by including a member of a talked-about fraternity, this additional study manipulated such norms by including a professor who was widely known to advocate group rights. This less subtle manipulation produced a significant backfiring effect when participants communicated to their other "friend" in a different context (consistent with Studies 1–3); however, when communicating in the presence of the professor, this backfiring effect was attenuated, and the moderation pattern was significant. Thus, the more obvious and recognizable the cue, the more likely the "immediate presence" moderating variable will exert its influence.

Controlling for the self-censorship cue manipulation did not substantially reduce the size of the relationship between attributions and disparaging communication intent (zero-order attribution-disparaging discussion,  $r = .27$ ,  $p < .001$ ; when controlling for the manipulation,  $r = .24$ ,  $p < .001$ ). Thus, although controlling for attributions reduced the predictive validity of the manipulation, the reverse was not true. Of importance, a Sobel  $z$  test indicated that this mediational pattern was not likely due to sampling error (Sobel  $z = 4.86$ ,  $p < .001$ ). These results are consistent with our expectation that attributional processes partially mediate the relationship between the self-censorship cue and the amount of disparaging communication. (Analyses performed on each study separately yielded an identical pattern of statistically significant mediation.)

### Comparing Reactance and Attributions As Mediators

Of course, although this mediational pattern is consistent with the attributional model, it does not rule out the possibility that more emotion-based processes such as reactance can account for the present effects. Thus, we analyzed Study 3 (the only study where we have a reactance measure directly available) to compare the mediation power of the two constructs. These analyses showed that, when considering Study 3 only, attributions about consensus were still a significant mediator of the self-censorship—disparaging communication relationship (zero-order backfiring effect size = .15; controlling for attributions = .07; Sobel's test  $p < .001$ ). These analyses also showed that reactance was a significant mediator of the same relationship (zero-order backfiring effect size = .15; controlling for reactance = .11; Sobel's test  $p = .023$ ), although descriptively reactance showed a mediation pattern that was less strong.

Given that both attributions and reactance showed an overall pattern of statistically-significant mediation (and that they were significantly correlated with each other;  $r = .41$ ,  $p < .001$ ), we performed some additional tests of their relative predictive power. First, we computed (via linear regression) the predictive power of each mediation measure on disparaging communication, while controlling for both the self-censorship cue and the other mediation measure (so, for reactance, we computed reactance's ability to predict disparaging communication while simultaneously entering attributions and the self-censorship cue as predictors). These results supported the contention that attributions about consensus were stronger mediators than reactance in the present set of results. In particular, attributions about consensus remained a statistically significant predictor of disparaging communication intent ( $\beta = .17$ ,  $p = .011$ ), but reactance was no longer a significant predictor ( $\beta = .10$ ,

$p = .108$ ). We then computed Sobel's tests while also controlling for both mediation variables simultaneously. These results yielded a significant Sobel's  $z$  test for attributions (while also entering in reactance;  $z = 2.45$ ,  $p = .014$ ). However, the Sobel's  $z$  test for reactance (while also entering in attributions) was not significant ( $z = 1.51$ ,  $p = .131$ ).

Although these tests do not suggest that reactance played no part in the current study (nor are the differences between the two mediators that large), they do provide evidence at the least that something about attributions above and beyond sheer reactance is playing a direct role in the key backfiring effect.

## GENERAL DISCUSSION

Taken together, the results of Studies 1 to 3 provide consistent evidence that norms compelling self-censorship can backfire and instead cause observers to ultimately communicate the censored belief. In all studies, a cue making salient the operation of a self-censorship norm designed to reduce disparaging communication instead increased the negativity of subsequent communication. Also, across all studies, this effect occurred in part because the salience of the self-censorship norm changed participants' attributions about why other persons communicated favorably about the fraternity in the first place. Participants were more likely to question the legitimacy of the shared positive opinion when the self-censorship norm was salient, thus leading them to ultimately communicate more disparagingly about the fraternity.

These results have multiple implications for our understanding of the potential effectiveness of the use of speech as a tool to fight negative group perceptions. We now turn our attention to these implications.

### Manufacturing Consensus and Its Consequences

At a broad level, consensus sometimes emerges in groups both small and large from the "bottom up," simply as a result of the dynamics of group interaction, without any apparent coercion (see Bourgeois, 2002; Conway, 2004; Conway & Schaller, 1998; Conway, Sexton, & Tweed, 2006; Crandall, 1988; Harton & Bourgeois, 2004; Opp, 1982; Schaller & Conway, 1999, 2001; Tweed & Conway, 2006). Other times, however, group consensus may emerge from the "top down" as a result of obvious social or environmental pressures (see, e.g., Berger & Heath, 2005; Berry, 1994; Cavalli-Sforza, 1993; Kitayama et al., 2006; Milgram, 1974; Vandello & Cohen, 1999). This susceptibility of popular opinion to social/environmental pressures makes it possible that powerful societal norms can simply manufacture consensus. If observers

believe that a powerful societal norm dictates something, then (all else being equal) those observers will obey. Thus, the activation of a societal norm encouraging the censorship of negative communication can potentially manufacture the appearance of a reduction in negative beliefs about groups very quickly.

The present scenario results suggest, however, that this kind of manufactured group consensus may be psychologically fragile. It is possible that people may sometimes comply with the self-censorship norm when it is salient and strong, thus producing less negative outward communications about groups; but, in a situation where the norm is either not salient or not strong enough to compel public assent, the superficial—and more positive—consensus collapses. Although of course we cannot speak directly from our data to what happens with real shared stereotypes in the real world, our results do interestingly parallel what often happens beyond the laboratory. Persons do not feel the same pressure of self-censorship norms at all times and all places. The present work suggests that this temporal variance in the norm's power has psychological consequences: Although people doubtless sometimes engage in self-censorship when they feel like they have to, it nevertheless (silently, but no less importantly) undermines their ability to personally believe the positive group communications they might hear. Thus, when they themselves do not feel directly compelled by the self-censorship norm, the norm produces a kind of rebound effect—actually causing more negative beliefs about the group to be communicated. In this less-public underground of communication, negative beliefs about groups may grow and flourish. Thus, like work on the impact of authority figure's commands in business contexts (Conway & Schaller, 2005), the present work suggests a consensus that is manufactured with a specific goal in mind (in this case, reducing negative communication) may have poor long-term prospects. As a tool for reducing the overall negativity of group perceptions, overtly making it salient to persons that they should censor negative communications may backfire.

After all, it is worth noting that the vast majority of even the most negatively stereotyped groups have many positive traits included in their shared stereotypes (see, e.g., Schaller et al., 2002). So what happens when people attempt to communicate those popular positive elements to others? The more salient self-censorship norms become, the less likely these positive elements will be believed. Thus, not without irony, it is possible that PC norms, like the institution of affirmative action (see Maio & Esses, 1998), the perception of a stereotyped group member's deviance from the stereotype (see Kunda & Oleson, 1997), or the communication of increasingly positive information relevant to minority group members (Collins, Biernat, & Eidelman, 2009),

may sometimes psychologically fertilize the very weed it seems designed to destroy. In the case of PC norms, it does so by making it more difficult for persons to believe—and thus ultimately communicate—any positive communications they hear about groups.

### Limitations and Caveats

This research is not without some limitations as well. First, these studies used fictitious scenarios. Although this is not a unique criticism of this work (e.g., most attribution and many stereotype studies have used some kind of scenario methodology) and low-external validity studies have many benefits (see, e.g., Mook, 1983), future research would do well to examine these processes in more real-world situations.

Second, some parts of the conceptual story were not entirely tidy. Cognitive load did not exert its expected influence on the backfiring effect in Study 1, and, although theoretically interpretable (see footnote 1), the immediate presence moderator did not exert its expected effect in Study 3. Still, taken as a whole package, we believe that the most plausible account of the present results is the attribution-based informational contamination account that inspired Studies 1 to 3. The present work provides a pattern of results that is almost entirely consistent with that attributional framework, and the mediation results suggest the plausibility of an attributional account. And, however interpreted, these results consistently demonstrate that, under some predictable circumstances, self-censorship norms may actually backfire.

### ENVOI

How do communication norms affect specific discussions about groups? The attributional model that inspired the present work encourages us to look beyond mere commonsense answers. Yes, self-censorship cues can sometimes lead to less negative communication. But there is more. Counterintuitively, those very things that are intended to discourage negative stereotypes can in fact end up leading to their increase. Just as the direct command of a business authority figure to engage in a consensual behavior can lead people to reject the behavior (Conway & Schaller, 2005), self-censorship cues can ultimately lead to more communication of the censored belief. These processes occur, in large part, because—despite intentions—these cues serve to divorce the positive communication of others from reality in the minds of the persons who may be potential carriers of that belief. Thus, the present approach encourages looking a little deeper at those psychological processes that determine exactly how people make attributions about

the group-relevant communications they hear. In doing so, it may help us better understand why social movements like political correctness do not appear to be entirely succeeding.

### ACKNOWLEDGMENTS

We thank Dov Cohen, Peter Darke, Darrin Lehman, Del Paulhus, and (especially) Mark Schaller for their invaluable help in theory and/or manuscript development.

### REFERENCES

- Alexander, M. G., Brewer, M., & Herrmann, R. K. (1999). Images and affect: A functional analysis of out-group stereotypes. *Journal of Personality and Social Psychology, 77*, 78–93.
- Arkin, R. M., Gabrenya, W. K., Appelman, A. S., & Cochran, S. T. (1979). Self-presentation, self-monitoring, and the self-serving bias in causal attribution. *Personality and Social Psychology Bulletin, 5*, 73–76.
- Baumgardner, A. H. (1991). Claiming depressive symptoms as a self-handicap: A protective self-presentation strategy. *Basic and Applied Social Psychology, 12*, 97–113.
- Berger, J. A., & Heath, C. (2005). Idea habitats: How the prevalence of environmental cues influences the success of ideas. *Cognitive Science, 29*, 195–221.
- Berry, J. W. (1994). Ecology of individualism and collectivism. In U. Kim, H. C. Triandis, C. Kagitcibasi, S.-C. Choi, & G. Yoon (Eds.), *Individualism and collectivism: Theory, method, and applications* (pp. 77–84). Thousand Oaks, CA: Sage.
- Bourgeois, M. (2002). Heritability of attitudes constrains dynamic social impact. *Personality and Social Psychology Bulletin, 28*, 1063–1072.
- Brehm, J. (1966). *Psychological reactance: A theory of freedom and control*. New York: Academic Press.
- Cavalli-Sforza, L. L. (1993). How are values transmitted? In M. Hechter, L. Nadel, & R. E. Michod (Eds.), *The origin of values* (pp. 305–317). New York: Aldine De Gruyter.
- Collins, E. C., Biernat, M., & Eidelman, S. (2009). Stereotypes and the communication of person impressions. *Journal of Experimental Social Psychology, 45*, 368–374.
- Conway, L. G., III. (2004). Social contagion of time perception. *Journal of Experimental Social Psychology, 40*, 113–120.
- Conway, L. G., III, & Gornick, L. J. (2009). [Why self-censorship norms fail – and succeed]. Unpublished raw data, The University of Montana.
- Conway, L. G., III, & Schaller, M. (1998). Methods for the measurement of consensual beliefs within groups. *Group Dynamics: Theory, Research, and Practice, 2*, 241–252.
- Conway, L. G., III, & Schaller, M. (2005). When authority's command backfire: Attributions about consensus and effects on deviant decision-making. *Journal of Personality and Social Psychology, 89*, 311–326.
- Conway, L. G., III, & Schaller, M. (2007). How communication shapes culture. In K. Fiedler (Ed.), *Frontiers of social psychology: Social communication* (pp. 107–127). New York: Psychology Press.
- Conway, L. G., III, Sexton, S. M., & Tweed, R. G. (2006). Collectivism and governmentally initiated restrictions: A cross-sectional and longitudinal analysis across nations and within a nation. *Journal of Cross-Cultural Psychology, 37*, 1–23.
- Conway, L. G., III, Thoemmes, F., Allison, A., Towgood, K. H., Wagner, M., Davey, K., et al. (2008). Two ways to be complex and why they matter: Implications for attitude strength and lying. *Journal of Personality and Social Psychology, 95*, 1029–1044.
- Crandall, C. S. (1988). Social contagion of binge eating. *Journal of Personality and Social Psychology, 55*, 588–598.
- Crawford, M. T., McConnell, A. R., Lewis, A. C., & Sherman, S. J. (2002). Reactance, compliance, and anticipated regret. *Journal of Experimental Social Psychology, 38*, 56–63.
- Drout, C. E., & Corsoro, C. L. (2003). Attitudes toward fraternity hazing among fraternity members, sorority members, and non-Greek students. *Social Behavior and Personality, 31*, 535–544.
- Eagly, A. H., & Wood, W. (1982). Inferred sex differences in status as a determinant of gender stereotypes about social influence. *Journal of Personality and Social Psychology, 43*, 915–928.
- Fein, S. (1996). Effects of suspicion on attributional thinking and the correspondence bias. *Journal of Personality and Social Psychology, 70*, 1164–1184.
- Fein, S., Hilton, J. L., & Miller, D. T. (1990). Suspicion of ulterior motivation and the correspondence bias. *Journal of Personality and Social Psychology, 58*, 753–764.
- Feldman, R. S., & Forrest, J. A., & Happ, B. R. (2002). Self-presentation and verbal deception: Do self-presenters lie more? *Basic and Applied Social Psychology, 24*, 163–170.
- Fosterling, F. (1989). Models of covariation and attribution: How do they relate to the analogy of analysis of variance? *Journal of Personality and Social Psychology, 57*, 615–625.
- Fuegen, K., & Brehm, J. W. (2004). The intensity of affect and resistance to social influence. In E. S. Knowles & J. A. Linn (Eds.), *Resistance and persuasion* (pp. 39–64). Mahwah, NJ: Erlbaum.
- Gilbert, D., & Hixon, J. (1991). The trouble of thinking: Activation and application of stereotypic beliefs. *Journal of Personality and Social Psychology, 27*, 26–47.
- Gurevitch, Z. D. (1985). The receiver's dilemma: Impressions formed in response to impression management. *Basic and Applied Social Psychology, 6*, 145–157.
- Harkins, S. G., & Petty, R. E. (1987). Information utility and the multiple source effect. *Journal of Personality and Social Psychology, 52*, 260–268.
- Harton, H. C., & Bourgeois, M. J. (2004). Cultural elements emerge from dynamic social impact. In M. Schaller & C. S. Crandall (Eds.), *The psychological foundations of culture* (pp. 41–76). Mahwah, NJ: Erlbaum.
- Hewstone, M., & Jaspars, J. (1987). Covariation and causal attribution: A logical model of the intuitive analysis of variance. *Journal of Personality and Social Psychology, 53*, 663–672.
- Indick, W. (2003). Individual variations in responses to a hypothetical moral dilemma. *Individual Differences Research, 1*, 29–38.
- Kelley, H. H. (1967). Attribution theory in social psychology. In D. Levine (Ed.), *Nebraska symposium on motivation* (Vol. 15, pp. 192–238). Lincoln: University of Nebraska Press.
- Kitayama, S., Ishii, K., Imada, T., Takemura, K., & Ramaswamy, J. (2006). Voluntary settlement and the spirit of independence: Evidence from Japan's "Northern Frontier". *Journal of Personality and Social Psychology, 91*, 369–384.
- Knowles, E. S., & Linn, J. A. (2004). The importance of resistance to persuasion. In E. S. Knowles & J. A. Linn (Eds.), *Resistance and persuasion* (pp. 3–9). Mahwah, NJ: Erlbaum.
- Kolditz, T. A., & Arkin, R. M. (1982). An impression management interpretation of the self-handicapping strategy. *Journal of Personality and Social Psychology, 43*, 492–502.
- Kunda, Z., & Oleson, K. C. (1997). When exceptions prove the rule: How extremity of deviance determines the impact of deviant examples on stereotype. *Journal of Personality and Social Psychology, 72*, 965–979.

- Maio, G. R., & Esses, V. M. (1998). The social consequences of Affirmative Action: Deleterious effects on perceptions of groups. *Personality and Social Psychology Bulletin*, *24*, 65–74.
- McArthur, L. A. (1972). The how and what of why: Some determinants and consequences of causal attribution. *Journal of Personality and Social Psychology*, *22*, 171–193.
- Miles, J. A., & Naumann, S. E. (2003). The effects of group affiliation on third parties' justice perceptions. *Current Research in Social Psychology*, *8*, 162–176.
- Milgram, S. (1974). *Obedience to authority*. New York: Harper and Row.
- Mook, D. G. (1983). In defense of external invalidity. *American Psychologist*, *38*, 379–387.
- Opp, K. D. (1982). The evolutionary emergence of norms. *British Journal of Social Psychology*, *21*, 139–149.
- Pryor, J. B., Rholes, W. S., Ruble, D. N., & Kriss, M. (1984). A developmental analysis of salience and discounting in social attribution. *Representative Research in Social Psychology*, *14*, 30–40.
- Rothbart, M., & Park, B. (1986). On the confirmability and discomfortability of trait stereotypes. *Journal of Personality and Social Psychology*, *50*, 131–142.
- Schaller, M., & Conway, L. G., III (1999). Influence of impression-management goals on the emerging contents of group stereotypes: Support for a social-evolutionary process. *Personality and Social Psychology Bulletin*, *25*, 819–833.
- Schaller, M., & Conway, L. G., III (2001). From cognition to culture: The origins of stereotypes that really matter. In G. B. Moskowitz (Ed.), *Cognitive social psychology: The Princeton symposium on the legacy and future of social cognition* (pp. 163–176). Mahwah, NJ: Erlbaum.
- Schaller, M., Conway, L. G., III, & Tanchuk, T. (2002). Selective pressures on the once and future contents of ethnic stereotypes: Effects of the “communicability” of traits. *Journal of Personality and Social Psychology*, *82*, 861–877.
- Tweed, R. G., & Conway, L. G., III (2006). Fundamental beliefs and coping across cultures. In P. T. P. Wong & L. C. J. Wong (Eds.), *Stress and coping: Multicultural & international perspectives* (pp. 133–153). New York: Kluwer Academic.
- Tyler, J. M., & Feldman, R. S. (2005). Deflecting threat to one's image: Dissembling personal information as a self-presentation strategy. *Basic and Applied Social Psychology*, *27*, 371–378.
- Vandello, J. A., & Cohen, D. (1999). Patterns of individualism and collectivism across the United States. *Journal of Personality and Social Psychology*, *77*, 279–292.
- Wegener, D. T., Clark, J. K., & Petty, R. E. (2006). Not all stereotyping is created equal: Differential consequences of thoughtful versus nonthoughtful stereotyping. *Journal of Personality and Social Psychology*, *90*, 42–59.
- Wikipedia. (2008). *Political correctness*. Retrieved December 9, 2008, from [http://en.wikipedia.org/wiki/Political\\_correctness](http://en.wikipedia.org/wiki/Political_correctness)
- Wilder, D. (1977). Perception of groups, size of opposition, and social influence. *Journal of Experimental Social Psychology*, *13*, 253–268.