

# When Authorities' Commands Backfire: Attributions About Consensus and Effects on Deviant Decision Making

Lucian Gideon Conway, III  
Indiana State University

Mark Schaller  
University of British Columbia

The perception of consensus typically produces conformity, but specific attributional circumstances may produce deviance instead. Ironically, the command of an authority figure may create one such circumstance. Participants were presented with scenarios in which they had to make a choice between 2 options. Prior to their decision, they observed others all making a single choice. In some conditions, this consensus occurred following an authority's explicit command to make that choice. Results revealed the hypothesized effect—the authority's command led participants to make deviant decisions—and revealed that this effect was moderated by the authority's continued presence, expertise, the target(s) of the command, and the ability of perceivers to use their cognitive resources.

*Keywords:* attributions about consensus, stereotypes, communication, popular opinion, consensus

Consensus not only defines popular opinion, it provides a basis for its persistence. When we observe others all expressing the same belief, endorsing the same product, or engaging in the same behavior, we tend to be more likely ourselves to express, endorse, or engage in that same thing (Crandall, 1988; Prentice & Miller, 1993; Schaller, 2001). This creates a sort of normative inertia: Consensually shared beliefs and behaviors tend to be self-replicating.

There are at least two sets of psychological processes underlying the influential power of consensus. One set of processes pertains to mere conformity. Mindful of the social rejection that often attends deviance, individuals strategically choose to go along with the crowd—at least in public—even if they do not privately endorse the popular opinion (Asch, 1956). But consensus inspires more than mere conformity. Others' actions are an important means through which knowledge about the world is gained; consensus serves as a sort of "social proof," attesting to the value of beliefs and behaviors that attract consensual endorsement (Cialdini, 1993; Hardin & Higgins, 1996; Matz & Wood, 2005). So, a second set of processes pertains to the effects of consensus on individuals' cognitions and inferences about the object of popular opinion. These inferential processes are the focus of the research reported here.

---

Lucian Gideon Conway, III, Department of Psychology, Indiana State University; Mark Schaller, Department of Psychology, University of British Columbia, British Columbia, Canada.

This research was funded by an Isaak Walton Killam doctoral fellowship awarded to Lucian Gideon Conway, III, and by Social Sciences and Humanities Research Council of Canada Research Grant 410-2001-1435 awarded to Mark Schaller. Two of the studies reported here were part of Lucian Gideon Conway, III's doctoral dissertation at the University of British Columbia. We thank Jennifer Casterline, Gary Evans, Arkisha Gathright, Molly Jameson, Rachael Lord, Margaret Ryan, and Rodney Scott for their assistance in data collection and coding as well as Kathrene R. Conway for generating the idea that turned into Study 4.

Correspondence concerning this article may be addressed to Lucian Gideon Conway, III, who is now at the Department of Psychology, University of Montana, Missoula, MT 59812. E-mail: luke.conway@umontana.edu

The conceptual foundation underlying this research represents a simple extension of familiar attributional logic. This simple analysis, however, yields novel hypotheses about circumstances under which popular opinion does and does not persist. We derive and test one particular hypothesis about the consequences of an authority figure's command: Because of its effects on the attributional process, an authority's command to express a particular opinion (an event that typically compels consensual obedience) may lead individuals to question the inherent value of a consensually held opinion, and so may compel them to reject the popular opinion. The five studies reported here test this hypothesis and reveal a predictable set of circumstances under which an authority's commands backfire in exactly this way. These studies also test an additional set of hypotheses, identifying variables that moderate this phenomenon.

## Effects of Perceived Consensus on Individuals' Attitudes and Opinions

There is abundant evidence that other's opinions—especially when they are perceived to represent a consensus—influence individuals' own perceptions and recollections of the world around them (Bless, Strack, & Walther, 2001). For instance, individuals express prejudices more strongly when they perceive that those prejudicial beliefs are normatively endorsed or shared more widely (Crandall, Eshleman, & O'Brien, 2002; Stangor, Sechrist, & Jost, 2001). This is not merely strategic self-presentation either; stereotypes spring to mind more easily when individuals perceive that others endorse those stereotypes (Sechrist & Stangor, 2001). Similar effects are witnessed on attitudes more widely. In general, if it is perceived that most others hold a particular attitude, then that attitude is more likely to be endorsed (Darke, Chaiken, Bohner, Einwiller, Erb, & Hazelwood, 1998; Eagly & Chaiken, 1993).

This influence of consensus information on individuals' own attitudes may reflect a semirational appraisal of the logical implications of consensus. If the vast majority of office workers use a particular software product, for instance, then it is reasonable to infer that it must be a pretty good product. The logical basis of this

popular-equals-good inference is sufficiently appealing, and the inference itself is made with sufficient frequency that, like many other inferential judgments (e.g., Gilbert, 1989), it may become a heuristic, that is, engaged quickly and used without conscious awareness.

This consensus heuristic, however, does not apply to all persons at all times. Sometimes people seem particularly impacted by popular opinion; but other times, they seem unaffected by it. Why? One set of factors that predicts the impact of popular opinion on perceivers can be understood in terms of attribution theory.

#### Attribution-Based Inferences About the Objects of Consensus and About Consensus Itself

Perceivers make attributions about the object of consensus—that thing that is popularly endorsed. This process has been studied a great deal, especially in the area of person perception (Försterling, 1989; Hewstone & Jaspars, 1987; Kelley, 1967; McArthur, 1972). As this research reveals, the perception of consensus limits the tendency for perceivers to draw conclusions about the qualities of those persons who participate in that consensus and instead directs perceivers to draw conclusions about things external to those persons—such as the object of consensus. (“That must be a great piece of software if everyone’s using it.”) But consensus information is not always equally informative. The influence of consensus on attributions—and on consequent beliefs—is moderated by many aspects of the immediate context (Alicke & Insko, 1984; Darke et al., 1998; Maheswaran & Chaiken, 1991; Wright, Luus, & Christie, 1990; Zuckerman & Feldman, 1984). For instance, the impact of consensus is substantial when one perceives that others’ agreement is because of independent evaluations of a particular attitude object but is undermined when one perceives that independent assessment to be contaminated by the operation of interpersonal influence (Harkins & Petty, 1987; Wilder, 1977).

It appears, therefore, that perceivers not only use consensus information to make attributions about objects of consensus but also make attributions about the consensus itself (see, e.g., Kenworthy & Miller, 2002; O’Laughlin & Malle, 2002). Just as we ask “Why is this person doing this thing?” we may also ask “Why is everyone doing the same thing?” Sometimes—especially when little conscious thought is devoted to the attribution process—the default answer to that question may implicate the inherent qualities of the object of that consensus. But other times, attention may be drawn to contextual features that offer alternative circumstantial explanations for the existence of consensus. Anything that makes individuals question the causes of consensus (and so less likely to take it at face value) may make them less likely to go along with it.

What factors are likely to change persons’ attributions about consensus? Group categorization appears to be one factor. It is more likely that popular in-group attitudes are attributed to features of objective reality, but popular out-group attitudes are attributed to rather different factors (Kenworthy & Miller, 2002), and so people are more likely to accept and conform to his or her own groups’ attitudes. Group categorization is only one of many variables that may impact attributions about consensus, and little research to date explores the ultimate implications of these attributions on subsequent endorsements of popular opinion. Indeed, still little is known about when popular beliefs within a group are accepted and when those same popular beliefs may instead inspire deviance. To help fill this gap, the present work explored addi-

tional implications of individuals’ attributions about consensus. Some of these implications are ironic, suggesting that variables that usually compel popular opinion may also—if they become the focus of attributional attention—sow the psychological seeds for its undoing. We focus on one such variable: the command of an authority figure.

#### Attributional Effects of Perceived Obedience: When an Authority’s Commands Backfire

When an authority figure tells people to do something, they typically do it (Blass, 1991; Milgram, 1974). The psychology of obedience, therefore, is easily harnessed as a means through which to create consensus. Dictatorial fiat—whether issued in a state-house, a corporation, or a classroom—can manufacture popular opinion.

The attributional analysis, however, implies that this sort of popular opinion may be fragile and unlikely to persist in the absence of continued coercion. There are attributional consequences to the perception of obedience. When perceivers believe that others’ actions are constrained by circumstances that demand obedience, they are less likely to attribute those actions to those others’ actual opinions (Fein, 1996; Fein, Hilton, & Miller, 1990; Pryor, Rholes, Ruble, & Kriss, 1984). Thus, an authority’s command may call into question (where otherwise there would not have been a question) the legitimacy of a perceived consensus. For example, office workers may all use a particular software product because they actually all think that it is a user-friendly and an effective tool, and they may encourage others to use it as well. But if a memo is then handed down from an office manager commanding them to use that product rather than other available alternatives, then they may question its popularity and, if given an opportunity, may actually be more likely to switch to an alternative product.

In a sense, the command of an authority figure is a double-edged sword: It creates consensus among people who feel compelled to obey the authority figure, but it also undermines the psychological power of consensus in the minds of others who perceive their actions. As a means of ensuring the persistence of public opinion, an authority’s commands may backfire by making consensus less compelling.

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

Of course, people may resist conformity pressures for a number of reasons that have nothing to do with attributional processes (Knowles & Linn, 2004). Perhaps the most prominent example is found in psychological reactance theory (e.g., Brehm, 1966; Crawford, McConnell, Lewis, & Sherman, 2002; Fuegen & Brehm, 2004). Real or perceived pressure from others to engage in a particular behavior often causes feelings of a loss of personal freedom to choose. Freedom of choice is a valued psychological commodity, and so reactance theory posits that one often will deviate from others’ expectations—both in belief and in action—in order to reassert one’s right to choose. Thus, like the attributional model, reactance theory also predicts that under some circumstances, an authority’s command 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 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.

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, including some hypotheses that cannot be explained by reactance (thus, testing the unique explanatory value of the attributional model). Supportive results, of course, cannot imply that people never deviate because of 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 in ways that reactance (and other affective) processes cannot.

### Necessary Conditions and Moderating Variables

Although an authority figure’s commands may backfire—by undermining the informational power of popular opinion and inspiring deviance instead—the attributional model does not suggest that this will occur in all circumstances. Indeed, the effect is likely to occur only under a predictable set of conditions, and, thus, a number of variables are likely to moderate it. The present research tested the moderating impact of several such variables.

One variable pertains to the continued presence or control of the authority figure. Regardless of individuals’ independent opinions, they are unlikely to deviate from an obedience-based consensus if they feel compelled to obey the command of an authority figure. The context in which those commands backfire should be limited to those in which individuals are, over time, released from such immediate coercion.

Another variable pertains to the perceived expertise of the authority figure. For some authority figures, their authority derives primarily from their superior power or status in a hierarchical system; they may not be perceived to have any special knowledge or expertise relevant to the commands they issue. Other authority figures, however, may be perceived to have some special expertise or insight that informs their commands. Conformity with an expert’s directives may not be attributed merely to blind obedience, and even if consensus is attributed to obedience, the opinions of an expert may have an independent influence on individuals’ attitudes and actions (Eagly, 1983). Authority commands are most likely to backfire when authority is based primarily on power not on expertise.

A third moderating variable pertains to the audience of the authority’s command. The attributional analysis implies that the command is likely to backfire only when it can logically account for the perceived consensus, such as when it is communicated to all persons who compose the consensus. A command directed just to one individual may inspire reactance-based deviance on the part of that one individual (Brehm, 1966) but cannot erode popular opinion through the attributional process described above.

Additional moderating variables pertain to the willingness or ability of observers to fully apply their cognitive resources. Attributions that lead to deviance may require more effortful cognition than simply using the tried-and-true consensus heuristic. Thus, the “cognitive load” experienced by perceivers may moderate the impact of attribution-related cues on deviance.

### Overview of the Present Research

The five studies reported below were designed to empirically test six hypotheses derived from the attributional analysis and rule out alternative explanations.

All five studies tested the central hypothesis that defines the basic backfiring phenomenon under inquiry here: Under conditions in which perceivers observe consensus in others’ actions, explicit commands from an authority figure lead to a decrease in perceivers’ own willingness to engage in those actions.

These studies also tested four additional hypotheses bearing on limiting conditions and moderating variables. Study 1 tested the hypothesis that the backfiring effect is moderated by the continued presence of the authority figure issuing the command (the effect occurs most strongly when the authority figure has no continuing presence). Study 2 tested the hypothesis that the effect is moderated by the perceived expertise of the authority figure (the effect occurs most strongly when the authority figure is perceived to be a nonexpert in the relevant domain). Study 3 tested the hypothesis that the effect is moderated by the nature of the audience to whom the command is communicated (the effect occurs most strongly when the command is directed to all individuals who compose the consensus). Study 5 tested the hypothesis that the effect is moderated by the availability of cognitive resources (the effect occurs most strongly when cognitive resources are maximally available).

In addition, Study 4 examined the possible impact of the nature of the command (hard vs. soft) and the leadership style of the authority figure (authoritarian vs. open-minded) on deviance. These variables may be expected to have an influence on emotion-related deviance, but less so on attribution-related deviance. By showing that these variables have minimal impact, Study 4 helps to rule out these alternative explanations and so more fully implicates an attributional account.

Within each study, measures were taken of perceivers’ attributions about the causes of consensus. This allowed us to test an additional hypothesis about mediating mechanisms: Within those specified conditions in which an authority’s command leads to deviance from consensus, this effect is expected to be mediated by attributions about the causes of consensus.

### Study 1

Participants were presented with detailed descriptions of a corporate decision-making scenario in which employees within a company were asked to decide which of two computer networking systems to purchase and install. Participants took the perspective of one of these employees, who was charged with the responsibility of making a final network decision. Within their role, participants witnessed all others unanimously agreeing that one network (“WobbleNet”) was the better choice. Participants were later asked to render their decision. They also completed measures designed to assess attributions.

Within this context, two variables were manipulated and crossed within a  $2 \times 2$  experimental design. One variable described the command of an authority figure: In one condition, participants witnessed the company president issue an explicit public command to endorse WobbleNet; in the other condition, the president issued no such command. The second variable described the continued coercive presence of the president: In one condition the president continued as the participants' boss; in the other condition, the president was replaced by a new boss prior to the participants' final decision.

Whether in the same-boss or the old-boss condition, the boss' command is expected to change participants' attributions about consensus from "it must be a good system" to "the boss told them to vote for it." However, these attributional processes are not always expected to guide behavioral decision making; sometimes, external pressure may override internal attributions. Thus, it was expected that, when the old boss was no longer present at the end of the story, his prior command to "pick WobbleNet" would backfire, resulting in an ironic deviance from both his command and the consensus. However, if the old boss remained throughout the story, then the attributional processes discussed here should be overridden by the compulsion to obey a powerful authority figure. Thus, the backfiring effect should be attenuated—or even reversed—in this condition.

## Method

*Participants.* Eighty University of British Columbia undergraduates participated for extra credit in introductory psychology courses. Participants were randomly assigned to read one of the four experimental conditions within the  $2 \times 2$  design.

*Decision-making scenario.* Scenarios began with a "cast of characters" that defined participants' role as a senior vice president within the corporation. The cast of characters also described the participants' boss, the company president ("President Whim"), as a man with a background in marketing who knew very little about computers. President Whim was also depicted as enigmatic and powerful, frequently crushing those who crossed him. Participants then read a detailed description of a specific corporate decision-making situation. Participants were members of a committee that was to decide which of two computer networking systems the company should purchase and install. The committee members would be meeting to discuss and vote on the options. Participants would be chairing the meeting and would not vote, at least not initially. The description also provided some limited initial information about the relative advantages and disadvantages of the two competing networking systems. One system ("WobbleNet") was described as a "tiny bit faster," but the other system ("NetHawk") was described as much more reliable.

*Manipulation of an authority figure's command.* The next segment of the scenario contained the command manipulation. This manipulation was contained within a speech made by President Whim to the committee at the outset of their meeting. In the command condition, President Whim stated, "It seems clear to me that WobbleNet is the better system, so I think we should go with that. I do not want any disagreement over this," and later reiterated that "I'll be very disappointed if people do not vote for WobbleNet." In the no-command condition, President Whim instead said, "I do not want to influence the vote, so I'm not even going to tell you what I think. I want each of you to vote exactly what you think," and later reiterated that "I'll be very disappointed if people do not vote what they really think." Regardless of condition, everyone at the meeting (eight committee members) proceeded to vote for WobbleNet.

*Manipulation of continued authority.* The final segment of the scenario described a phone call that the participant's character received a few days later. It was President Whim calling. In the same-boss condition, Whim

reminded participants that he is still their boss and that they will be answering to him. In the new-boss condition, Whim informed participants that he was leaving the company to take a different job and that they would henceforth be answering to a new boss. In all conditions, participants later discovered that the committee's previous decision (the unanimous vote supporting WobbleNet) had been made moot and that participants themselves—as senior vice president and chair of the Computer Networking Committee—should choose whether to purchase WobbleNet or NetHawk. (In the same-boss condition, this new plan was delivered by President Whim; in the new-boss condition, it was delivered by their new boss, "Kirk Caprice.")

*Manipulation checks.* Immediately after reading the scenario, participants responded to questions assessing their memory for key elements of the story. In addition to a question about what system the committee voted for, two additional questions ensured the checked attention to information that varied across conditions: what system President Whim favored (if any), and who the participants' boss was at the end of the story. Each question was followed by three possible responses, and participants were asked to circle the correct response. Results revealed highly accurate recollection for these key elements of information (percentages correct ranged from 96% to 98%).

*Attributions about consensus.* Two methods were used to assess participants' attributions about the observed consensus. One method was open-ended. Participants responded in writing to the question, "Why do you think the committee members all voted for WobbleNet?" Responses were coded by Lucian Gideon Conway, III (while blind to experimental condition) on a 9-point bipolar scale, anchored at the high end by attributions to the inherent qualities of the networking systems (e.g., committee members believed WobbleNet to be the better system) and at the low end by attributions to social influence (e.g., pressure from President Whim or from other committee members). Ratings near the middle of the scale indicated some combination of these different attributions. (A second trained coder—who was blind to experimental condition and also unaware of the conceptual hypotheses—rated responses from a subset of participants. Interrater reliability was high,  $r = .96$ .) In addition to this open-ended response measure, participants also completed a set of three ratings (on 9-point scales) more explicitly assessing different possible explanations for the unanimous vote favoring WobbleNet. One rating assessed the inference that the consensus occurred "because WobbleNet is, in reality, the better system." Another rating assessed the inference that the consensus occurred "because [committee members] each really believed that WobbleNet was the better system." A third rating assessed the inference that the consensus occurred "because of the influence of President Whim." The four different attribution measures were highly intercorrelated, and so a composite measure was created: After reverse scoring the final rating scale measure (which was negatively correlated with the other three measures), scores on all four measures were converted to  $z$  scores, and the mean was computed. The resulting attribution index ( $\alpha = .89$ ) indicates the degree to which participants attributed the observed consensus to inherent qualities of the attitude object (indicated by more highly positive scores) or to constraints imposed by the social situation (indicated by lower scores on the index).

*Eventual endorsement.* Participants' eventual tendency to go along with—or deviate from—popular opinion was assessed by their rated responses to two questions assessing an inclination to choose one or the other networking system: "What is the likelihood that you would choose WobbleNet as the system your company would buy?" and "What is the likelihood that you would choose NetHawk as the system your company would buy?" Responses were recorded on 9-point rating scales. These responses were highly negatively correlated ( $r = -.82$ ), and so a composite measure was created. The rated preference for "NetHawk" was subtracted from the rated preference for "WobbleNet," and a constant of 5 was added to retain the original 9-point scale. Higher scores on this endorsement index indicate an increased likelihood to make a decision that is

consistent with the observed consensus; lower scores indicate a greater tendency to make a decision that deviates from popular opinion.<sup>1,2</sup>

## Results

Table 1 summarizes the mean responses on attribution and endorsement measures within each of the four conditions of the  $2 \times 2$  experimental design. These results are consistent with the conceptual hypotheses.

**Attributions about consensus.** We anticipated that participants would be more likely to attribute consensus to inherent qualities of WobbleNet when there were no obvious constraints on the committee's unanimous vote (the no-command condition) but that they would attribute consensus to some form of social influence in the command condition. This effect should not be moderated by the continued-authority manipulation. This pattern is exactly what occurred. On the attribution index, there was a main effect only of the command manipulation,  $F(1, 75) = 82.10, p < .001$ . The continued-authority variable exerted no main effect, nor did it moderate the effect of command on these attributions.

**Eventual endorsement.** We predicted that the attributional impact of the command variable would consequently influence participants' eventual endorsement of WobbleNet but that this effect on decision making would not occur if they were still under the coercive power of the authority figure. Results on the endorsement index indicated exactly this effect. There was a main effect of the command manipulation,  $F(1, 76) = 4.96, p = .029$ . However, this main effect was qualified by a Command  $\times$  Continued Authority interaction,  $F(1, 76) = 3.98, p = .050$ . The means in Table 1 reveal that the authority's command backfired—leading participants to endorse a position that deviated from popular opinion—within the new-boss condition,  $t(38) = 2.97$ , one-tailed,  $p = .003$ , but that this backfiring effect did not occur within the same-boss condition,  $t(38) < .20, p > .80$ .

**Evidence of mediation.** Additional analyses were used to test more directly whether the predicted effect of the authority's command on deviant decision making (within the new-boss condition) was mediated by the attributional impact of the command. Among participants in the new-boss condition, there was a negative zero-order correlation between the command manipulation (coded as 0 = no command, 1 = command) and the endorsement index ( $r = -.43$ ). However, when scores on the attribution index were controlled for, in a partial correlation analysis, this negative correlation disappeared entirely (and was somewhat reversed;  $pr = .26$ ). However, the correlation between the attribution index and endorsement index remained strong, even after controlling for the command manipulation (zero-order  $r = .71$ ;  $pr = .65$ ). These

results are consistent with an attributional explanation for the backfiring effect of the authority's command. A Sobel (1982) test indicated that this mediational pattern is unlikely to arise from sampling error alone ( $z = 4.32, p < .001$ ).

**Secondary mediation analyses.** We expected that, although participants' attributions would still be impacted by the president's command in the same-boss condition, these attributions would be overridden by the desire for compliance with the command of their boss. (As reported above, this exact pattern occurred). Because (as expected) attributional processes were still impacted by the command in the same-boss condition, it is possible that participants' attributions played an indirect role in that condition by suppressing a positive effect of the command on eventual endorsement. In fact, this exact pattern occurred. Among participants in the same-boss condition, there was no zero-order correlation between the command manipulation and the endorsement index ( $r = -.00$ ). However, when scores on the attribution index were controlled for, this correlation became positive ( $pr = .23$ ). A Sobel (1982) test indicated that this suppression pattern is unlikely to arise from sampling error alone ( $z = 1.99, p = .047$ ).

## Discussion

These results reveal that when an authority figure commands people to act in a consensual manner, then that command can ultimately lead individuals to act in ways that deviate from that consensus. The results also reveal the operation of an attributional process underlying this effect. One reason why an authority's command can ultimately lead to deviant decision making is that it changes individuals' attributions about the consensus they observe in others. In the absence of an authority's command, participants inferred that consensus reflected the inherent qualities of the object

<sup>1</sup> In addition to the measures described here in detail, participants in all studies also completed three other measures that, in different ways, assessed opinions about the two networking systems. One was a crude forced-choice measure in which participants simply circled which of the two networks they would choose. Two additional rating-scale measures assessed the favorability of participants' overall impressions of each of the two networks. Effects of independent variables on these measures were descriptively consistent with those on the endorsement index, but these effects were weak and typically did not meet most individuals' subjective standards for statistical significance (i.e.,  $ps > .05$ ). For this reason, coupled with the fact that these measures are either methodologically cruder or less immediately relevant to actual decision making, analyses on these additional measures are not presented here. Furthermore, participants in all studies completed an additional open-ended question about their attributions for their own behavior. This measurement was largely exploratory and not relevant to the present conceptual purpose and, thus, is not presented here.

<sup>2</sup> Dependent measures were presented to participants in two different orders. Some participants received the free-response questionnaire first, followed by the rating scales; others received the rating scale questionnaire first. (Within both the free-response and rating scale questionnaires, the endorsement items were always presented prior to the attribution items. In the text, items are presented in a theoretical, rather than in a chronological, order.) Preliminary analyses revealed no questionnaire order effects on either the attribution or endorsement measures, nor did order moderate any of the predicted effects of other manipulated variables on these measures. This variable is not discussed further. (In subsequent studies, all participants received measures in the same order, with the free-response questionnaire first.)

Table 1  
*Interactive Effects of Authorities' Command and Continued Presence of Authority on Attributions and Endorsement of the Object of Consensus in Study 1*

Measure	Same boss		New boss	
	No command	Command	No command	Command
Attribution	0.51	-0.61	0.71	-0.64
Endorsement	4.75	4.55	5.80	2.15

of consensus; in contrast, a command led them to attribute the consensus to the power of social influence instead.

A subsequent effect on deviant decision making, however, occurs only when individuals feel free of the coercive power of that authority figure. When individuals remain under that person's power, the fragile consensus holds. Just because private attributions are made does not mean that they are always acted on. It is clear, however, that the attributional impact is real and has consequences. Sometimes these consequences are obvious: If freed from the perceived necessity to obey the authority figure, then attributions do guide behavior, and deviance may occur. Sometimes these consequences are less obvious. Even under conditions in which participants experienced the boss' continued authority, there was evidence that the attributional process suppressed the powerful compulsion to comply with the command of that authority figure.

Thus, it is clear that authority figures' commands to engage in consensual behavior can lead to deviant decision making, that this effect occurs for predictable reasons, and that there are predictable limits on this effect. Study 2 tested an additional hypothesis about these limits.

## Study 2

In Study 1, participants were explicitly informed that the authority figure was not particularly knowledgeable about the topic on which consensus emerged. Effects of an authority's command may be quite different, however, when that authority figure is perceived to be an expert on this topic.

There are two conceptually distinct ways in which expertise may plausibly exert this moderating effect. One possibility is that the authority's expertise may influence perceivers' attributions about the causes of consensus; when the authority is an expert, the consensus may be attributed to the inherent qualities of the object (e.g., WobbleNet) endorsed by this expert rather than to mere obedience. If so, then an expertise manipulation should exert a main effect on attributions and should interact with a command manipulation in predicting participants' own eventual endorsement. A second possibility is that because expert status is such a compelling persuasion cue in its own right (Eagly, 1983), the authority's expertise may exert a direct effect on participants' attitudes and actions without necessarily having any impact on the attributional process. If so, then no main effect of an expertise manipulation should be observed on attributions, but it may, nonetheless, interact with a command manipulation in predicting participants' endorsement.

To test these possibilities, participants in Study 2 were presented a decision scenario based on the new-boss condition from Study 1. Within this scenario, a command manipulation was introduced, just as in Study 1. In addition, an expertise manipulation was introduced: The authority figure was either an expert or a nonexpert on the relevant topic. These manipulations were crossed in a  $2 \times 2$  factorial design.

## Method

**Participants.** Ninety-two undergraduate students from the University of British Columbia participated for extra credit. Participants were randomly assigned to one of four conditions within the  $2 \times 2$  experimental design.

**Procedures and manipulations.** The scenario presented to participants was identical to that of the new-boss condition from Study 1. Within the context of this scenario, the command manipulation was also the same as in Study 1: In the command condition, President Whim commanded committee members to vote for WobbleNet; in the no-command condition, he instructed them simply to vote according to their own thoughts. A second manipulation was introduced at the outset of the scenario (when introducing the "cast of characters"): Participants in the nonexpert condition read that President Whim had obtained a master's degree in marketing and "does not know anything about computers"; participants in the expert condition read that he had obtained a master's degree in computer networking and "knows a whole lot about computers." (Manipulation checks indicated that participants attended to and remembered key elements of the story, including information about President Whim's expertise; percentages of accurate responding ranged from 95% to 100%.)

**Attribution and endorsement measures.** All measures were identical to those used in Study 1. (Interrater  $r = .98$  for coding of the free-response attribution measure.) An attribution index and an endorsement index were computed in the same way as in Study 1.

## Results

Table 2 presents mean responses on attribution and endorsement measures within each condition of the  $2 \times 2$  experimental design. These results are consistent with the hypotheses.

**Attributions about consensus.** As in Study 1, there was a strong main effect of the command manipulation,  $F(1, 88) = 88.88, p < .001$ . Compared with participants in the no-command condition, those in the command condition were less likely to attribute consensus to the inherent qualities of the computer networks and instead tended to attribute consensus to social influence. The expertise manipulation had neither a meaningful main effect nor a moderating effect on these attributions ( $F_s < 1.8, p_s > .190$ ).

**Eventual endorsement.** There was a main effect of the command manipulation on participants' eventual endorsement,  $F(1, 88) = 13.88, p < .001$ . However, this effect was qualified by an interaction between the command and expertise manipulations that approached but did not attain significance,  $F(1, 88) = 3.42, p = .068$ . When President Whim was a nonexpert, there was a sizable difference between the no-command and command conditions in participants' endorsement of WobbleNet, thus replicating the effect found in Study 1,  $t(44) = 4.24, p < .001$ ; the difference was smaller and nonsignificant when President Whim was a computer expert,  $t(44) = 1.21, p = .233$ .

**Evidence of mediation.** Additional correlations and partial correlations were computed for participants in the nonexpert condition (for whom the backfiring effect was expected and found). Consistent with expectations, the impact of the command manipulation on the endorsement index was substantially reduced when

Table 2  
*Interactive Effects of Authorities' Command and Expertise on Attributions and Endorsement of the Object of Consensus in Study 2*

Measure	Expert		Nonexpert	
	No command	Command	No command	Command
Attribution	0.70	-0.53	0.52	-0.69
Endorsement	5.04	3.52	5.35	0.83

controlling for scores on the attribution index (zero-order  $r = -.55$ ,  $p < .001$ ;  $pr = -.20$ ,  $p = .196$ ). Although the relation between the attribution and endorsement indices was unexpectedly reduced when controlling for the command manipulation (zero-order  $r = .61$ ,  $p < .001$ ;  $pr = .35$ ,  $p = .020$ ), these results, nonetheless, imply that the effect of command on endorsement was partially mediated by attributions. Indeed, a Sobel (1982) test indicated that (as in Study 1) this mediational pattern is unlikely to result from sampling error alone ( $z = 2.30$ ,  $p < .021$ ).

*Secondary mediation analyses.* Parallel mediation analyses were also performed for the expert condition. As in Study 1, these secondary analyses revealed that attributional processes suppressed a positive effect of the command on compliance. Among participants in the expert condition, there was a negative zero-order correlation between the command manipulation and the endorsement index ( $r = -.18$ ). However, when scores on the attribution index were controlled for, this correlation became strongly positive ( $pr = .38$ ). A Sobel (1982) test indicated that this suppression pattern is unlikely to result from sampling error alone ( $z = 4.19$ ,  $p < .001$ ).

### Discussion

Study 2 replicated and extended the results of Study 1. When a nonexpert authority figure commanded others to vote for WobbleNet—and they did so unanimously—participants were more likely later to make a decision that deviated from the observed consensus. This effect was substantially mediated by attributions about why the consensus existed. This backfiring effect was reduced, however, when the authority figure was an expert on the relevant topic. In this expert condition, participants' eventual decisions tended to conform to the consensus whether or not the expert commanded them to.

It is interesting to note that the authority figure's expertise did not have any meaningful impact on attributions but affected endorsement anyway. This pattern of results indicates that the authority's expertise exerted a more direct effect on eventual decision making by providing another independent cue that influences individual attitudes and actions. This is not inconsistent with the attributional model. It is hardly surprising that the expertise of the authority figure did not change the impact of an authority's command on attributions; even if observers believe that an authority is an expert, they are still likely aware of the potential for his command to induce direct compliance in other people, thus making those observers more likely to think that the consensus emerged because of the command. Of course, that does not mean that the authority's expertise is entirely bankrupt: It may still partially offset its attributional impact by providing a direct independent influence on participants' perceptions. Mediation analyses within the expert condition suggest that this very thing happened. When removing the effect of attributions, the authority command exerted a positive effect on compliance.

Indeed, Studies 1 and 2 both suggest moderating variables that operate independently of attributions and thus serve to offset the impact of those attributions. In Study 1, the continued authority of the boss (despite having no impact on attributions) counteracted the attributional impact that his command had on deviance. Similarly, in Study 2, the expertise of the boss (again despite having no impact on attributions) partially counteracted the attributional impact of his command on deviance. Thus, these variables moderate

the backfiring effect through direct impact on compliance or individual perceptions of the computer systems without impacting attributions. This is underscored by the fact that in both studies, attributions about consensus suppressed a positive effect of the command on compliance in the moderating conditions. If the moderating variables had changed the impact of the command on attributions (i.e., either making them obsolete or reversing their effect), then this suppression effect would have been improbable.

However, other variables may moderate the backfiring effects of an authority's commands on decision making by directly impacting attributions. Study 3 was designed to test just such a moderating variable. Inquiry into this new moderating variable also tested the viability of a plausible alternative process—psychological reactance—through which an authority's command may inspire deviant decision making.

### Study 3

When individuals feel that freedom of choice is restricted, they respond by finding ways to reestablish that decisional freedom. This reactance phenomenon can lead individuals to disobey directives or otherwise deviate from the expectations of others (Brehm, 1966). Although a reactance process does not explain results on the attribution measures taken in Studies 1 and 2, it may offer some partial explanation for the finding that an authority's command can backfire and ultimately lead to deviance. (Note that in Study 2, the effect of the command manipulation on endorsement was only partially mediated by attributions.) Study 3 was designed to test more directly the extent to which reactance may provide a plausible account for the effects of an authority's command on eventual deviation from consensus.

This design included a variable that, logically, should not affect the reactance process but that should influence the attributional process. This variable concerns the target(s) of the authority's command. In one condition, the authority's command was clearly communicated to the entire set of people who later voted unanimously to endorse WobbleNet. In another condition, this command was directed solely to participants. Because reactance is predicated on individuals' perceived threats to their own decisional freedom, this manipulation should not affect any tendency for reactance processes to guide participants' ultimate endorsement of WobbleNet. However, this manipulation should influence an attributional process. (Obedience is a plausible explanation for consensus—and so may lead to deviant decision making—only if the authority's command is directed toward everyone, but not if it is directed toward a single individual.) If this manipulation moderates the effects of an authority's command on endorsement, then it uniquely implicates the operation of an attributional process rather than a reactance process.

### Method

*Participants.* Participants were 91 undergraduate students from Indiana State University. They were randomly assigned to one of four conditions within a  $2 \times 2$  (Command  $\times$  Target of Instructions) experimental design.

*Procedures and manipulations.* Participants were presented with a scenario that, in most respects, replicated the nonexpert condition from Study 2. As in previous studies, a command manipulation was introduced through the instructions that President Whim offered at the outset of the committee meeting: In the command condition, President Whim issued a

command to vote for WobbleNet; in the no-command condition, he issued no such command. In addition, wording varied across scenarios to create a target-of-instructions manipulation: Participants in the group-target condition read that these instructions were spoken to the entire committee (as in Studies 1 and 2); those in the individual-target condition read that President Whim pulled the participant's character aside and ensured that only their character heard the instructions. It was emphasized the no one else on the committee heard these special instructions.

Manipulation check results indicated that participants generally remembered most key parts of the story with near-perfect accuracy, as in previous studies, but only 78% of participants correctly identified the target of President Whim's instructions. Because of this low percentage, we opted to present here analyses only for those persons who made correct identifications on this manipulation check. Including the entire sample of participants revealed results very similar to those presented below.<sup>3</sup>

**Measures.** Measures of attribution and endorsement were identical to those used in Studies 1 and 2. (Interrater  $r = 1.00$  for coding of the free-response attribution measure.) An attribution index and an endorsement index were computed in the same way as in previous studies. In addition, three items (with responses on 9-point rating scales) were included to assess feelings of reactance (e.g., "To what degree did President Whim's discussion at the committee meeting make you feel as if he was trying to take away your freedom to do exactly as you wished?"). Responses to these three items were highly correlated, and so a composite reactance index was computed as the mean rating across the items ( $\alpha = .79$ ).

## Results

Table 3 presents mean responses on attribution, reactance, and endorsement measures within each condition of the  $2 \times 2$  experimental design. These results are consistent with the conceptual hypotheses.

**Attributions about consensus.** Both independent variables had main effects on the attribution index (both  $ps < .008$ ), but these main effects were qualified by the predicted  $2 \times 2$  interaction,  $F(1, 67) = 26.43, p < .001$ . When participants believed that the entire group was the target of President Whim's instructions, there was a strong effect of the command manipulation (this replicates the effect observed in Studies 1 and 2). However, when participants believed that they were the sole target of these instructions, the effect of the command manipulation difference was negligible. Moreover, the results indicated that it was only in the command/group-target condition that consensus was attributed to social influence rather than to the inherent qualities of the networking systems.

**Reactance.** Results for the reactance index were similarly consistent with expectations. Participants in the command condition experienced substantially more reactance than did those in the

no-command condition,  $F(1, 67) = 48.20, p < .001$ . The target-of-instructions manipulation had no main effect, nor did it moderate the effects of the command condition on the reactance index (both  $ps > .310$ ).

**Eventual endorsement.** The command manipulation had essentially no main effect on the endorsement index ( $p = .603$ ), and there was only a weak main effect for the target-of-instructions manipulation ( $p = .100$ ). However, consistent with the results on the attribution measure—and with expectations—there was an interactive effect of these two variables on the endorsement index,  $F(1, 67) = 4.61, p = .035$ . When President Whim's command to endorse WobbleNet was directed at the whole group, the command backfired, as in Studies 1 and 2, leading participants eventually to be less likely to endorse that product,  $t(29) = 1.07$ , one-tailed,  $p = .151$ . (Although falling short of significance by conventional standards, this result clearly replicates the pattern from Studies 1 and 2, and so we have a strong a priori confidence that it is not because of sampling error.) No such effect of the command manipulation was observed when these instructions were directed to participants alone; in fact, the pattern of means was actually reversed within the individual-target condition, an effect that approached significance,  $t(38) = -1.96, p = .058$ .

**Evidence of mediation.** Additional analyses focused just on those participants for whom the backfiring effect occurred—those in the group-target condition. Results revealed that the negative zero-order correlation between the command manipulation and endorsement index entirely disappeared (and was actually reversed) when controlling for scores on the attribution index (zero-order  $r = -.21; pr = .34$ ). The strongly positive relation between the attribution and consensus indices remained essentially unchanged when controlling for the command manipulation (zero-order  $r = .42; pr = .49$ ). These results are consistent with the hypothesis that attributions about consensus mediated the relation between the command manipulation and eventual endorsement. A Sobel (1982) test indicated that this mediational pattern is unlikely to emerge from sampling error alone ( $z = 2.85, p = .004$ ).

In contrast, there was very little evidence that reactance mediated that relation: When scores on the reactance index were controlled, there was only a trivial reduction in the size of the relationship between the command manipulation and the endorsement index (zero-order  $r = -.21; pr = -.10$ ; Sobel  $z = 0.42, p = .674$ ).

**Secondary mediation analyses.** In Studies 1 and 2, the moderating variable showed no impact on attributions, instead exerting an independent and direct impact on decision making. However, consistent with expectations, in Study 3, the target-of-instruction manipulation moderated the impact of the command on attributions about consensus: In the group-target condition, the authori-

Table 3  
*Interactive Effects of Authorities' Command and Target of Command on Attributions, Reactance, and Endorsement of the Object of Consensus in Study 3*

Measure	Individual target		Group target	
	No command	Command	No command	Command
Attribution	0.31	0.17	0.65	-1.00
Reactance	3.93	6.30	3.30	6.50
Endorsement	3.11	6.05	3.64	1.85

<sup>3</sup> For the whole sample (i.e., including even those participants who failed to correctly identify the target of the command), the key Target  $\times$  Command interaction was still in evidence, interaction  $F(1, 87) = 7.32, p = .008$ . (When only including persons who incorrectly identified the target, the same interaction effect is nonexistent [ $F < .001, p = .991$ ], suggesting that persons who misidentified the target appear to be simply adding random noise.) In addition, mediation analyses on the whole sample suggested that, although attributions about consensus mediated the key backfiring effect (Sobel  $z = 3.74, p < .001$ ), reactance played little of a mediational role (Sobel  $z = 0.99, p = .320$ ). Thus, the key results reported in the text hold even if all the participants in Study 3 are included.



ty's command had a large influence, whereas in the individual-target condition, the command had minimal influence. Because (unlike the moderating conditions in Studies 1 and 2) attributional processes were not expected to be in operation in the individual-target condition in Study 3, they were not expected to mediate or suppress effects in that condition. Secondary analyses generally supported this contention. Among participants in the individual-target condition, there was a positive zero-order correlation between the command manipulation and the endorsement index ( $r = .30$ ); when scores on the attribution index were controlled for, this correlation became more positive ( $pr = .49$ ), but a Sobel (1982) test revealed that this modest suppression pattern was likely because of sampling error ( $z = 0.66, p = .510$ ).

### Discussion

These results identify another (expected) limiting condition on the backfiring effect, and in doing so, they more fully implicate attributional processes rather than reactance processes as the explanation for this effect. The command of an authority figure led participants eventually to make a decision that deviated both from the command and from the consensus they observed, but this effect occurred only when the command was communicated to all persons who composed that consensus. No such effect occurred when the command was directed at participants alone. In this latter condition, the command aroused reactance—just as it did when it was directed at the whole group—but it did not substantially undermine the typical tendency to infer inherent quality in the object of consensus. Mediation analyses further revealed that when an authority's command backfires by leading others to deviate from popular opinion, it does so as a causal consequent of its effects on attributional inference. Secondary mediation analyses also revealed that, consistent with the expectation that attributional processes were not in operation in the individual-target condition, no significant suppression or mediation effects occurred in that condition.

### Study 4

Study 4 provided further tests of reactance as an alternative explanation for the backfiring effect by introducing two variables that should moderate the impact of reactance processes but that should have no meaningful impact on the attributional process. These two variables referred to the style and demeanor of the authority figure.

Not all authority figures are alike. Some (such as those depicted in Studies 1–3) are authoritarian and dictatorial; others are more congenial and open-minded. Similarly, not all commands are alike. Some commands (such as those used in Studies 1–3) are like hard, blunt implements that demand compliance; other commands are softer, gentler, and encourage compliance through more subtle means. What is the likely impact of these different styles of leadership on the kind of deviance documented in Studies 1 through 3? The answer differs, depending on the particular process that allegedly underlies this deviance.

From our attributional perspective, there is no reason to expect any real impact of these variables. From this perspective, the backfiring phenomenon depends on attributions about consensus, and there is no necessary reason why differences in style should result in differences in consensus. Rigid authoritarians are not

necessarily better at building consensus than people who adopt a more congenial and open-minded managerial style. Nor does a blunt, hard command necessarily exert more impact on consensus than a softer, more subtle approach to social influence. To the degree that a leader is perceived as potentially influential, it is irrelevant whether or not that influence comes from an authoritarian style or from some other source. Thus, from an attributional point of view, the kind of deviance documented in the preceding studies is unlikely to be moderated by variables pertaining to leadership style.

In contrast, these kinds of variables should influence reactance. Subtle and nonthreatening tactics of influence do not induce the same reactance processes as overt and threatening tactics (Fuegen & Brehm, 2004). Thus, if a reactance process underlies the tendency for an authority's command to backfire, then this tendency should be reduced when the authority figure adopts a more open-minded style or delivers commands in a softer, subtler manner.

### Method

*Participants.* Participants were 91 undergraduate students from Indiana State University. They were randomly assigned to one of six conditions within a  $3$  (command: none, soft, or hard)  $\times 2$  (leader style: open or authoritarian) experimental design.

*Procedures and manipulations.* Participants were presented with a scenario that, in most respects, replicated the group-target condition from Study 3. As in previous studies, a command manipulation was introduced through the instructions that President Whim offered at the outset of the committee meeting. In the hard-command condition, President Whim issued a command to vote for WobbleNet; in the no-command condition, he issued no such command. A further command manipulation was also added in which Whim gave his preference for WobbleNet but said that he did not want to influence the vote. So in this soft-command condition, participants were aware that everyone knew their boss' preference, but they were not explicitly told to vote that way.

In addition, wording varied across scenarios to create a leader style manipulation: Participants in the authoritarian boss condition read that (as in Studies 1–3) President Whim was an enigmatic boss who was prepared to crush those who go against his will, whereas participants in the open-boss condition read that,

Whim's most noticeable personality trait is that he is very kind. Everyone knows that they can go to Whim for advice, help, or just to talk. He works very well in groups and is a very thoughtful, warm-hearted leader. People under him know that while working on a project, Whim works with his subordinates, not against them. People do not fully understand what drives his kindness, but they are thankful for it just the same. Another characteristic of this gentle man is that he tends to listen openly to those who cross him.

Manipulation check results indicated that participants generally remembered most key parts of the story with a high degree of accuracy (87%, 96%, and 92% retention), as in previous studies. Initial analyses excluding participants who made incorrect identifications on any of the manipulation check items revealed results very similar to those of the entire sample, both descriptively and inferentially; results on the entire sample are presented below. An additional manipulation check question revealed that the leader style manipulation had its intended effect: Participants rated the authoritarian boss as far less "nice" ( $M = 3.02$ ) than the open boss ( $M = 5.53$ ),  $t(88) = 7.50, p < .001$ .

*Measures.* Measures of attribution and endorsement were identical to those used in Studies 1–3. An attribution index and an endorsement index were computed in the same way as in previous studies.

Table 4  
*Effects of Authorities' Command and Leader Style on Attributions, Reactance, and Endorsement of the Object of Consensus in Study 4*

Measure	Authoritarian leader			Open leader		
	No command	Soft command	Hard command	No command	Soft command	Hard command
Attribution	0.76	-0.40	-0.51	0.86	-0.22	-0.45
Reactance	4.04	6.09	6.73	3.11	4.11	7.00
Endorsement	4.00	2.07	1.69	7.33	1.33	3.07

## Results

Table 4 presents mean responses on attribution, reactance, and endorsement measures within each of the six conditions of the 3 × 2 experimental design.

*Attributions about consensus.* A main effect emerged for the command manipulation on the attribution index,  $F(2, 85) = 35.78$ ,  $p < .001$ . As can be seen in Table 4, participants were more likely to attribute the observed consensus to qualities inherent in the computer systems when no command was issued than when either a soft or hard command was issued. Thus, no matter the format through which the president's wishes were made known, his command had a large impact on participants' attributions. No main effect of leader style, nor a Leader Style × Command interaction, emerged ( $F_s < 0.7$ ).

*Reactance.* A main effect emerged for the command manipulation on the reactance index,  $F(2, 85) = 23.92$ ,  $p < .001$ . Participants were especially likely to experience reactance in the hard-command condition, less likely in the soft-command condition, and especially unlikely in the no-command condition. Participants were also more likely to experience reactance when their leader was authoritarian than open,  $F(1, 85) = 5.10$ ,  $p = .026$ . Furthermore, an interaction that approached significance emerged between leader style and command, such that participants experienced far more reactance when an authoritarian leader gave a soft command than when an open leader did so,  $F(1, 85) = 2.80$ ,  $p = .067$ . These results reveal that, although leader style had no impact on attributions about consensus, it did have the expected impact on reactance (see Fuegen & Brehm, 2004).

*Eventual endorsement.* A main effect emerged for the command manipulation on eventual endorsement,  $F(2, 85) = 7.05$ ,  $p = .001$ . As can be seen in Table 4, the president's endorsement of WobbleNet, whether given in a hard or soft manner, reduced the likelihood that participants would also give their endorsement: hard-command/no-command comparison,  $t(58) = 3.41$ ,  $p < .001$ ; soft-command/no-command comparison,  $t(59) = 2.63$ ,  $p = .011$ . The hard- and soft-command conditions did not differ from each other,  $t(59) = -0.67$ ,  $p = .506$ .

Participants were somewhat less likely to endorse WobbleNet if the president showed an authoritarian style of leadership, although this effect was nonsignificant,  $F(1, 85) = 2.08$ ,  $p = .153$ . The Leader Style × Command interaction was also nonsignificant ( $F < 1.60$ ,  $p > .20$ ).

*Evidence of mediation.* Because the hard-command and soft-command conditions yielded similar results on the attribution index, they were combined for mediation analyses and both dummy-coded as 1 (with the no-command condition coded as 0). Results revealed that the negative zero-order correlation between

the command manipulation and endorsement index almost entirely disappeared when controlling for scores on the attribution index (zero-order  $r = -.36$ ,  $p < .001$ ;  $pr = -.02$ ,  $p = .817$ ). The strongly positive relation between the attribution and consensus indices remained largely unchanged when controlling for the command manipulation (zero-order  $r = .52$ ,  $p < .001$ ;  $pr = .40$ ,  $p < .001$ ). These results are consistent with the hypothesis that attributions about consensus mediated the relation between the command manipulation and eventual endorsement. A Sobel (1982) test indicated that this mediational pattern is unlikely to result from sampling error alone ( $z = 4.53$ ,  $p < .001$ ).

In contrast, there was not much evidence that reactance mediated that relation: When scores on the reactance index were controlled, there was only a trivial reduction in the size of the relationship between the command manipulation and the endorsement index (zero-order  $r = -.36$ ,  $p < .001$ ;  $pr = -.26$ ,  $p = .013$ ).<sup>4</sup> Further analyses suggested that participants' perceptions of the president's niceness also did not mediate the impact of the command manipulation on endorsement (zero-order  $r = -.36$ ,  $p < .001$ ;  $pr = -.36$ ,  $p = .001$ ).

## Discussion

These results suggest that the backfiring effect is not dependent on an authoritarian boss, nor is it dependent on the command itself, indicating a strong desire for compliance. Indeed, even when an especially open-minded and congenial boss expressed his preference in a nondemanding way, that preference had a notable effect on participants' eventual deviance. These results are inconsistent

<sup>4</sup> More focused analyses suggested initially that reactance may have had a mediational impact when comparing only the hard-command with the control condition (effect size  $r = .32$ ; when controlling for reactance,  $r = .05$ ; Sobel test  $z = 2.09$ ,  $p = .037$ ). However, this effect appears to be accounted for in part by reactance's incidental overlap in Study 4 with attributions about consensus. A key component of mediation is the ability of the mediator to predict the dependent variable while controlling for the independent variable (see Baron & Kenny, 1986). When both proposed mediators are entered in simultaneously, with the command manipulation as predictors of endorsement, attribution remains strong and significant ( $\beta = .61$ ,  $p < .001$ ), whereas reactance becomes nonsignificant ( $\beta = .18$ ,  $p = .279$ ). Furthermore, focused comparisons between the soft-command and control conditions yielded little evidence that reactance mediated the backfiring effect (Sobel's  $z = 1.05$ ,  $p = .300$ ), whereas attributions clearly did provide mediation (Sobel's  $z = 3.30$ ,  $p < .001$ ). Thus, considered along with evidence from Studies 3 and 5, the cumulative evidence presented here points to an attributional (and not a reactance) explanation.

with a reactance approach and are perhaps intuitively surprising, but they are not surprising from an attributional point of view. If the backfiring effect hinges on attributions about consensus, then anything that implies that consensus emerged on account of social influence will facilitate deviance. These attributions do not depend on leadership style: It does not matter how gentle and kindly an opinion is offered—nor how gentle and kind the leader who offered it—as long as perceivers believe the utterance could have had social influence.

Not only are the patterns of means inconsistent with a reactance process—and consistent with an attributional process—so, too, are the results of mediation analyses. In summary, these results further cast doubt on a reactance explanation for the backfiring effect and further support the hypothesized attributional process.

### Study 5

Although evidence suggests that the key effect is largely driven by the proposed attributional processes, it yet remains unclear exactly how those processes operate. Is the path from consensus-relevant attributional cues to ultimate decisions a noneffortful, automatic process that requires little in the way of cognitive resources? Or, does the link from attributions about consensus to decision making occur through effortful thinking and thus require substantial cognitive resources in order for attribution-relevant cues to exert their effects? Study 5 was designed to address these questions.

Previous research on attributions about individuals' behaviors suggests that the resources required to make and apply such attributions depend on the nature of the cues involved. When the cues are extremely salient, specific, or accessible, using those cues in making attributions does not require a lot of cognitive resources. Under other circumstances, however, using those cues in making attributions does require substantial cognitive resources (Trope & Gaunt, 2000). Although not directly mapping onto this previous work, the discounting attributions driving the key effect in Studies 1–4 are based on an authority figure's command that does not appear to be overly salient, specific, or accessible. Thus, the backfiring effect may be dependent on the availability of sufficient cognitive resources. In order to test this hypothesis, half the participants in Study 5 were presented with a manipulation designed to reduce access to cognitive resources.

There are two conceptually distinct ways in which cognitive load may plausibly exert this moderating effect. One possibility is that cognitive load may directly interfere with persons' abilities to recognize the attributional importance of relevant cues. If so, then the load manipulation should interact with authority command to predict both eventual endorsement and attributions. However, it is also plausible that cognitive load does not interfere with the recognition of attribution-relevant cues per se but rather with the (temporally later) application of those cues to decision making and attitude judgments. If so, then no main or interaction effect of the load manipulation should be observed on attributions, but it may, nonetheless, interact with a command manipulation in predicting participants' endorsement.

### Method

**Participants.** Participants were 202 undergraduate students from Indiana State University. They were randomly assigned to one of four conditions within a 2 × 2 (Command × Cognitive Load) experimental design.

**Procedures and manipulations.** Participants were presented with a scenario that replicated the group-target condition from Study 3. As in previous studies, a command manipulation was introduced through the instructions that President Whim offered at the outset of the committee meeting: In the command condition, President Whim issued a command to vote for WobbleNet; in the no-command condition, he issued no such command. (These commands were "hard" commands used in Study 4.) Participants were then freed from the president's authority near the end of the story.

In addition, participants were randomly assigned to either a load or no-load condition, which was contained in a set of written instructions on the first page of their experimental packet. Participants in the load condition were instructed to remember an eight-digit number as they read and answered questions about the story; participants in the no-load condition were not.

Manipulation check results indicated that participants generally remembered most key parts of the story with a high degree of accuracy—88%, 99%, and 96%. However, to ensure that the results of Study 5 cannot be accounted for by a differential likelihood of persons in the load condition to make errors pertaining to the story, initial analyses, while excluding those persons who incorrectly identified any aspect of the story, were performed. These results were very similar to those of the entire sample; results on the entire sample are presented below.

**Measures.** Measures of attribution and endorsement were identical to those used in Studies 1–4. An attribution index and an endorsement index were computed in the same way as in previous studies. In addition, as a check on the cognitive load manipulation, after completion of all the key dependent and mediation variables, participants were asked to recall the number. The vast majority of participants recalled most or all of the number. (Analyses excluding participants who did not appear to make an effort to remember the number yielded results virtually identical to those reported below.)

Furthermore, because the cognitive load manipulation may also affect participants' mood independent of any impact on cognitive resources, a mood questionnaire was also given to participants. This questionnaire consisted of five bipolar items similar to those used in previous research (e.g., Conway, 2004). A "mood" composite was computed by averaging these five items ( $\alpha = .77$ ).

### Results

Table 5 presents mean responses on attribution and endorsement measures within each condition of the 2 × 2 experimental design. These results are consistent with the hypotheses.

**Attributions, reactance, and mood.** Results of a 2 × 2 ANOVA revealed a main effect of the command manipulation on attributions about consensus: Compared with participants in the no-command condition, those in the command condition were less likely to attribute consensus to the inherent qualities of the computer networks and instead tended to attribute consensus to social

Table 5  
*Interactive Effects of Authorities' Command and Cognitive Load on Attributions, Reactance, and Endorsement of the Object of Consensus in Study 5*

Measure	No load		Load	
	No command	Command	No command	Command
Attribution	0.65	−0.65	0.60	−0.51
Reactance	3.23	6.04	3.59	5.28
Endorsement	5.64	1.92	4.59	4.40

influence,  $F(1, 195) = 195.83, p < .001$ . The cognitive load manipulation had neither a meaningful main effect nor a moderating effect on these attributions ( $F_s < 1.3, p_s > .260$ ).

A similar pattern emerged for reactance, yielding a main effect of the command manipulation,  $F(1, 197) = 62.45, p < .001$ , and no main effect for cognitive load ( $F < 0.5$ ). The interaction effect, although not large (see Table 5), approached significance,  $F(1, 197) = 3.87, p = .051$ .

The cognitive load manipulation appeared to have no bearing on the mood of participants: Participants were in virtually the same mood whether they were in the load ( $M = 4.61$ ) or no-load ( $M = 4.76$ ) condition ( $p > .310$ ). Similarly, the interaction between load and mood was nonsignificant ( $p > .15$ ).

*Eventual endorsement.* A  $2 \times 2$  ANOVA revealed a main effect for command  $F(1, 197) = 8.57, p = .004$ . This effect was qualified by the expected Cognitive Load  $\times$  Command interaction,  $F(1, 197) = 7.00, p = .009$ . As in Studies 1–4, when participants were not under cognitive load, they endorsed WobbleNet less when the president commanded them to choose it ( $M = 1.92$ ) than when he did not do so ( $M = 5.64$ ),  $t(93) = 4.00, p < .001$ . However, this backfiring effect of the president's command was virtually erased when participants were under cognitive load (command  $M = 4.40$ ; no command  $M = 4.59$ ),  $t(93) = 0.20, p = .843$ . (No main effect for load was evident [ $F < 1.2$ ].)<sup>5</sup>

*Evidence of mediation.* Additional analyses focused just on those participants for whom the backfiring effect occurred—those in the no-load condition. Results revealed that the negative zero-order correlation between the command manipulation and endorsement index disappeared (and was slightly reversed) entirely when controlling for scores on the attribution index (zero-order  $r = -.37, p < .001$ ;  $pr = .12, p = .245$ ). The strongly positive relation between the attribution and consensus indices remained essentially unchanged when controlling for the command manipulation (zero-order  $r = .59, p < .001$ ;  $pr = .51, p < .001$ ). These results are consistent with the hypothesis that attributions about consensus mediated the relation between the command manipulation and eventual endorsement. A Sobel (1982) test indicated that this mediational pattern is unlikely to emerge from sampling error alone ( $z = 4.96, p < .001$ ).

Controlling for reactance did not substantially reduce the size of the relationship between the command manipulation and the WobbleNet endorsement score (zero-order  $r = -.39, p < .001$ ; controlling for the reactance composite,  $pr = -.30, p = .003$ ).<sup>6</sup>

*Secondary mediation analyses.* Parallel mediation analyses were also performed within the load condition. As in Studies 1 and 2 (in which command also impacted attributions in the “moderating” conditions), these secondary analyses revealed that attributional processes suppressed a positive effect of the command on compliance. Among participants in the load condition, there was virtually no zero-order correlation between the command manipulation and the endorsement index ( $r = -.01$ ). However, when scores on the attribution index were controlled for, this correlation became strongly positive ( $pr = .49$ ). A Sobel (1982) test indicated that this suppression pattern is unlikely to emerge from sampling error alone ( $z = 6.30, p < .001$ ).

## Discussion

The results of Study 5 revealed that the impact of an authority figure's prior command on the endorsement of a consensual behavior is dependent on the availability of cognitive resources.

When participants were deprived of their resources, the backfiring effect of an authority's command virtually disappeared. This effect is not merely the result of a general impact on attention to the story: The results held (and were still statistically significant) even when only looking at those participants who correctly recalled all key aspects of the story.

What causes this moderating effect of cognitive load? The present study offers some clues. It is noteworthy that, unlike for the endorsement index, there was no pronounced effect of cognitive load on the attribution index. This suggests that load did not interfere with participants' ability to recognize the attributional import of the president's command but rather interfered with their ability to apply this attribution-relevant information to their decision making. (This contention is also supported by mediation

<sup>5</sup> Across all five studies, participants completed the Personal Need for Structure questionnaire (PNS; Neberg & Newsom, 1993; Thompson, Naccarato, Parker, & Moskowitz, 2001), which in part assesses persons' unwillingness to apply their cognitive resources in novel situations. Previous research reveals that individuals with higher PNS are less likely to take circumstantial constraints into account when drawing inferences about the actions of individuals and groups (Moskowitz, 1993; Schaller, Boyd, Yohannes, & O'Brien, 1995). To assess the degree that PNS moderated the backfiring effect of an authority's command on eventual endorsement, we (a) pooled the data from all five studies; (b) selected only those conditions that were identical across studies (i.e., an authoritarian leader gives a hard command) in which the backfiring effect was expected and found; (c) converted the command manipulation and PNS to  $z$  scores; (d) created a Command  $\times$  PNS interaction term; and (e) entered the command, PNS, and interaction terms into a simultaneous regression analyses predicting eventual endorsement. This analysis provided modest support for the hypothesis. The interaction between the authority command and PNS was small but significant ( $\beta = .12, p = .033$ ), and inspection of means of high- versus low-PNS persons revealed that the difference between command and no-command conditions was larger for low-PNS persons (compared with high-PNS persons). This result, like Study 5, is consistent with the idea that the key backfiring effect is in part dependent on the availability of and/or willingness to use cognitive resources. However, the moderating effect itself appeared to be moderated by the leader style and hard- and soft-command distinctions used in Study 4; when both soft commands and open leaders are included in the pooled analyses, the effect is still in the right direction, but the effect size diminishes ( $\beta = .05, p = .308$ ). Thus, the PNS analyses provided the same results when using the exact same comparisons as those found in Study 5 (providing a direct conceptual replication of that study) but suggest that the moderating effect of PNS occurs only at the extreme ends of the authoritarian leader style. The analyses also indirectly suggest that the moderating effect of cognitive load may only occur when authoritarian leaders issue hard commands, despite the fact that soft commands from open leaders also produce the backfiring effect.

<sup>6</sup> Across all five studies, additional mediation analyses were performed using only the measure of attributions that was not confounded with either measures of the president's command or participants' own perceptions of the two systems: “because [the committee members] each really believed that WobbleNet was the better system.” These analyses revealed the exact same mediational pattern across all five studies: The strong relationship between the command manipulation and eventual endorsement was substantially reduced when controlling for this more focused measure of attributions about consensus, but the relationship between attributions and endorsement remained strong when controlling for the command manipulation. In all five studies, this mediational pattern was statistically significant using a Sobel test ( $z_s > 2.98, p_s < .004$ ). Thus, the mediational analyses reported appear to reflect more than a simple bias in measurement; even when using a measure that only focuses on participants' views of the relationship between the observed consensus and committee members' actual beliefs, the same mediational pattern is evident.

analyses; for load participants, attributions about consensus were suppressing a positive effect of the command on compliance).

In some ways, Study 5 is reminiscent of research revealing the effects of cognitive load on attributional judgments of individual behaviors (Gilbert, Krull, & Pelham, 1988; Gilbert, Pelham, & Krull, 1988). For example, cognitively distracted persons in one study were less likely to take situational factors into account when making attributional judgments about why a woman appeared nervous. These cognitively busy participants recognized the situational constraint; they simply did not take it into account in ultimate judgments about the woman's nervousness (Gilbert, Pelham, & Krull, 1988, Experiment 1). One may view the present work as illustrating the same process in attributions about consensus and its consequences. Cognitively busy participants in our work recognized the situational constraint relevant to consensus but were unable to use that information when making decisions about the object of that consensus.

Alternatively, the present work may also be viewed as indirectly suggesting an additional "step" in the observation-to-decision model. Participants in our study made attributions about group behavior and then applied those attributions to a decision-making task. Thus, unlike in the studies by Gilbert, Pelham, and Krull (1988), the moderating impact of cognitive load observed here may occur not because of a direct effect of load on attributional processes per se but rather from the effect load has on a step farther along the decision-making chain. Given the potentially complicated nature of the relationships between cognitive load, recognition, and application (see, e.g., Gilbert & Hixon, 1991), future research would do well to more fully disentangle exactly how load exerts its impact on the backfiring effect.

Whatever its exact cause, however, it is clear that persons need their full complement of cognitive resources for the backfiring impact of an authority's command to occur robustly. If people do not have those resources, then they will be more likely to just go along with the crowd—and the command. This work thus complements previous theory and research, suggesting that when persons are unable or unwilling to think deeply, they are more susceptible to social influence (e.g., Kruglanski, Webster, & Klem, 1993; Richter & Kruglanski, 2003; Shestowsky, Wegener, & Fabrigar, 1998).

### General Discussion

The five studies reported here suggest that when an authority figure commands people to behave in a consensual manner, the command itself can ultimately lead individuals to deviate from the consensus. Thus, ironically, an authority's command can backfire—leading to the ultimate erosion of the popular opinion the command was intended to inspire.

These five studies also suggest that this erosion of popular opinion occurs, in large part, because the authority's command changes observers' attributions about why popular opinion exists in the first place. In the absence of an authority's command, observers assumed that popular opinion reflected favorably on the object of the opinion. However, when the command was present, observers attributed the existence of the popular opinion to the power of social influence instead.

As a consequence of this attributional process, there are clear and predictable boundaries for when the backfiring effect will occur. The observation of these moderating factors not only more clearly implicates attribution processes as an explanatory cause of

the backfiring effect but also highlights the ability of the underlying theoretical perspective to generate novel hypotheses and empirical discoveries (see Conway & Schaller, 2002).

### *Deviance From Authority: Beyond Reactance*

There has been a recent resurgence of interest in why people sometimes resist compliance attempts by other individuals (see, e.g., Knowles & Linn, 2004). Some of this work identifies factors that influence deviance beyond those considered in the traditional reactance model. For example, one factor that seems to moderate when people deviate from versus comply with a command is anticipated regret: Persons are much more likely to comply with a command when they think about the potential regret they may feel if their decision goes awry (Crawford et al., 2002). Although the primary goal of the present research was to evaluate deviance from consensus, the present results also have implications for rejecting the overt compliance attempts of other individuals—specifically authority figures. (Note that when participants deviated from the consensus in the present work, they also always deviated from an authority's command.) Thus, this research complements other theoretical approaches to compliance and deviance and suggests that this broad domain of inquiry may be fruitfully informed by a more thorough application of attributional principles.

Although our studies have focused on the conditions in which an authority's command may backfire, this research does not undermine the long-established contention that people feel strongly compelled to obey the commands of authority figures. When controlling for attributional processes in the present set of studies (by either statistical mediation analyses or experimental manipulation), an authority's command almost always induced compliance. However, our results suggest that, although an authority's command may indeed induce compliance, it may also change observers' subsequent attributions of others' compliance. As a result of this attributional influence, in situations in which authority figures are trying to replicate popular opinion, the command of an authority figure may actually cause more individual deviance than compliance. In addition, the present work also identifies exactly what sort of circumstances are especially likely to lead to deviance from an authority's command. We consider some of these circumstances more fully below.

### *The Rocky Road From Observed Consensus to Individual Opinion*

Within the context of impression formation, the causal path from the observation of behavior to the inference of inherent qualities has been aptly characterized as a rocky road (Jones, 1979). Consensus information is commonly used by observers to help navigate that rocky road so as to arrive at reasonably accurate inferences about the persons they perceive (Försterling, 1989; McArthur, 1972). As the present results revealed, perceivers not only make attributions on the basis of consensus information but also make attributions about the causes of consensus itself, and these attributions influence individual beliefs. Often these inferences are straightforward: Consensus about some object of opinion seems to indicate something about the inherent qualities of that object, and so individual opinions about that object conform to popular opinion. But things are not always so straightforward; the road from observed consensus to individual opinion can be quite rocky as well. Predictable aspects of the social context (e.g.,

evidence of social influence) are treated as extenuating circumstances, disrupting the easy tendency to merely replicate popular opinion. Moreover, additional variables moderate this tendency to consider extenuating circumstances and so complicate further the path from the observation of popular opinion to the formation of individual opinion.

Like other recent work on attributional processes, the present work highlights how complicated attribution can be. For example, the multiple inference model (Reeder, Vonk, Ronk, Ham, & Lawrence, 2004) suggests that, upon witnessing a particular behavior, observers make inferences relevant to multiple motives and then integrate those inferences into an overall impression. Because the present work was not designed to distinguish different motives from each other, it cannot tease apart the explanations derived from the multiple inference model, suspicion model (Fein, 1996), or more traditional attribution theory models (see Reeder et al., 2004, for a discussion of these models). However, the present work is consistent with the general picture that emerges from recent multiple inferences work of humans as attributionally complex processors, suggesting that people potentially form attributions at multiple levels (group and individual) and that the resulting judgments are impacted by multiple aspects of the situation. Thus, although not focusing on simultaneous inferences about multiple motives, the present work does similarly highlight the great complexity of online attributional processing.

### *The Persistence and Erosion of Popular Opinion*

Of course, the opinions of one person rarely matter very much. It is the opinions of the masses that really impact the course of societies. For this and other reasons, a psychological understanding of consensually shared beliefs is important to psychologists (see Conway & Schaller, 1998; Schaller & Conway, 2001). It is worth considering, then, that the present work not only has implications for how consensus impacts a single individual's opinion but also has implications for our understanding of the persistence and erosion of opinions that are widely shared.

Popular TV shows, soft drinks, religious systems, political groups—indeed, virtually every normative belief or behavior that has ever existed—cannot become or remain popular unless they continually “stick” in the mind or behaviors of an audience (see Gladwell, 2000). But this only begs the question: What sorts of things impact the ability of a belief or behavior to endure in the minds of large populations of people? Previous work has identified numerous predictive factors, including those that pertain to the match between a belief and a local ecology (e.g., Berger & Heath, 2005; Berry, 1994; Cohen, 1998; Conway, Ryder, Tweed, & Sokol, 2001; Insko et al., 1980; Tweed & Conway, in press) and those that pertain to the processes of interpersonal communication (e.g., Cavalli-Sforza, 1993; Schaller, 2001; Schaller & Conway, 1999, 2001; Schaller, Conway, & Tanchuk, 2002).

The present attributional framework complements this previous work by helping fill in the considerable gaps that still exist in the understanding of the emergence and persistence of popular beliefs (Cialdini & Trost, 1998; Nielsen & Miller, 1997). One basic message that underlies much previous research and theory on the persistence of popular beliefs is that, because of the power of consensus on individual minds, popular beliefs tend to be very durable. However, the attributional approach suggests that this influence of consensus itself should not be constant across all times and all places. Indeed, the influence of consensus rests on a

sometimes-implicit attributional judgment that consensus reflects favorably on the belief everyone professes. If obvious constraints—things that appear to have pushed the consensus into being—are made salient to us, then consensus is robbed of its power. It will not stick.

### *The Durability of Manufactured Consensus*

Sometimes, consensus arises spontaneously via communication (Colarelli, 1998; Conway, 2004; Hardin & Higgins, 1996; Opp, 1982). Other times, however, consensus occurs because it is demanded. The command of an authority figure inspires obedience, and so can be a very efficient means of manufacturing consensus, at least in the short term. But how durable is that consensus? How resistant is it to individual deviance?

The results from our studies revealed a set of variables that influence the durability of this sort of manufactured consensus. It is certainly more durable if the authority figure persists to exert control and influence over others. It is also more durable if that authority figure is perceived to have expertise in relevant domains. It is also more durable if it is less obvious to observers that the consensus may be manufactured by obedience. If commands are issued covertly—whispered to individuals one at a time rather than broadcast loudly to entire groups—the resulting consensus is likely to be attributed to genuine beliefs rather than to the operation of social influence, and individuals will likely to conform to that consensus on their own accord. Finally, these results indicate that a consensus manufactured by obedience is more likely to endure if those individuals who perceive the consensus do not have the cognitive resources available to think about things very deeply. If people are willing and able to think hard about consensus-relevant information, then manufactured consensus is more likely to collapse. However, if people do not process such information with much effort, they are especially likely to simply “go along with the crowd,” even when the crowd is engaging in an obviously manufactured consensus.

### *How Stable Are Attributional Inferences Over Time?*

Of course, to more fully understand how attributions about consensus influence the persistence of popular opinion, the stability of such consensus-related attributional inferences over time will have to be explored. Is the attributional inference (“An authority figure caused this apparent consensus”) itself highly resistant to change, even if subsequent information may call into question the validity of the original inference (e.g., if perceivers are made aware that the persons composing the original consensus really did think WobbleNet was the better product after all)?

The stability of these inferences over time can have important implications for their influence on popular opinion. If these attributions are fleeting, then their impact on the persistence of popular opinion may have little lasting impact. It would be like throwing a pebble in the ocean of popular opinion: It may cause a temporary and measurable fluctuation, but the next moment, the ocean would look almost exactly as it did before. However, if these attributions about consensus are fairly intractable, then even the most coercively manufactured consensus may be quite fragile under certain circumstances. A different metaphor would be necessary: Such intractable attributions would then operate like tiny flaws in a plane of glass, lurking unnoticed until a little stress (e.g., the leader who had manufactured the consensus is removed from authority)

applied in just the right spot may lead a seemingly durable consensus to suddenly shatter. This would suggest that as a long-term strategy, manufacturing consensus is a double-edged sword: Commanding consensus may work in the short term, but it may lead to a sudden complete collapse of that consensus later.

Whether attributionally induced deviance ends up being more like a pebble in an ocean or flaws on a plane of glass cannot be gleaned directly from the present work because we did not test how stable these attributions are over time. It is likely that the answer is complicated and that, as with other kinds of knowledge structures, the mutability of attributions about consensus is dependent on a host of other psychological factors (see, e.g., Wegener, Petty, Smoak, & Fabrigar, 2004). The exploration of this question, and its implications, is left to future research.

### Conclusion

Because of the powerful psychological effect of consensus, popular opinion has a tendency to self-replicate. Our research suggests that this normative inertia is dependent, in part, on observers' perceptions about why that consensus exists in the first place. When observers sense that popular opinion has been artificially manufactured, they are more likely to deviate from that consensus. Although straightforward on the surface, this simple extension of attribution theory has novel and ironic consequences for the persistence and erosion of popular opinion. Our research explored just one of these consequences: The predictable tendency for an authority's command to backfire. Many other potential consequences await discovery.

### References

- Alicke, M. D., & Insko, C. A. (1984). Sampling of similar and dissimilar comparison persons: Objects as a function of the generality of attribution goals. *Journal of Personality and Social Psychology, 46*, 763–777.
- Asch, S. E. (1956). Studies of independence and conformity: A minority of one against a unanimous majority. *Psychological Monographs, 70*, 416.
- Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173–1182.
- 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.
- Blass, T. (1991). Understanding behavior in the Milgram obedience experiment: The role of personality, situations, and their interactions. *Journal of Personality and Social Psychology, 60*, 398–413.
- Bless, H., Strack, F., & Walther, E. (2001). Memory as a target of social influence? Memory distortions as a function of social influence and metacognitive knowledge. In J. P. Forgas & K. D. Williams (Eds.), *Social influence: Direct and indirect processes* (pp. 167–183). Philadelphia: Psychology Press.
- 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.
- Cialdini, R. B. (1993). *Influence: Science and practice* (3rd ed.). New York: HarperCollins.
- Cialdini, R. B., & Trost, M. R. (1998). Social influence: Social norms, conformity, and compliance. In D. T. Gilbert & S. T. Fiske (Eds.), *The handbook of social psychology, Vol. 2* (4th ed., pp. 151–192). Boston: McGraw-Hill.
- Cohen, D. (1998). Culture, social organization, and patterns of violence. *Journal of Personality and Social Psychology, 75*, 408–419.
- Colarelli, S. M. (1998). Psychological interventions in organizations: An evolutionary perspective. *American Psychologist, 53*, 1044–1056.
- Conway, L. G., III. (2004). Social contagion of time perception. *Journal of Experimental Social Psychology, 40*, 113–120.
- Conway, L. G., III, Ryder, A. G., Tweed, R. G., & Sokol, B. W. (2001). Intranational cultural variation: Exploring further implications of collectivism within the United States. *Journal of Cross-Cultural Psychology, 32*, 681–697.
- 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. (2002). On the verifiability of evolutionary psychological theories: An analysis of the psychology of scientific persuasion. *Personality and Social Psychology Review, 6*, 152–166.
- Crandall, C. S. (1988). Social contagion of binge eating. *Journal of Personality and Social Psychology, 55*, 588–598.
- Crandall, C. S., Eshleman, A., & O'Brien, L. (2002). Social norms and the expression and suppression of prejudice: The struggle for internalization. *Journal of Personality and Social Psychology, 82*, 359–378.
- 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.
- Darke, P. R., Chaiken, S., Bohner, G., Einwiller, S., Erb, H., & Hazelwood, J. D. (1998). Accuracy motivation, consensus information, and the law of large numbers: Effects on attitude judgment in the absence of argumentation. *Personality and Social Psychology Bulletin, 24*, 1205–1215.
- Eagly, A. H. (1983). *Who says so? The processing of communicator cues in persuasion*. Paper presented at the meeting of the Eastern Psychological Association, Philadelphia, PA.
- Eagly, A. H., & Chaiken, S. (1993). *The psychology of attitudes*. Fort Worth, TX: Harcourt Brace.
- 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.
- Försterling, 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. T. (1989). Thinking lightly about others: Automatic components of the social inference process. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 189–211). New York: Guilford Press.
- Gilbert, D. T., & Hixon, J. (1991). The trouble of thinking: Activation and application of stereotypic beliefs. *Journal of Personality and Social Psychology, 60*, 509–517.
- Gilbert, D. T., Krull, D. S., & Pelham, B. W. (1988). Of thoughts unspoken: Social inference and the self-regulation of behavior. *Journal of Personality and Social Psychology, 55*, 685–694.
- Gilbert, D. T., Pelham, B. W., & Krull, D. S. (1988). On cognitive busyness: When person perceivers meet persons perceived. *Journal of Personality and Social Psychology, 54*, 733–740.
- Gladwell, M. (2000). *The tipping point: How little things make a big difference*. Boston: Little, Brown.
- Hardin, C., & Higgins, E. T. (1996). Shared reality: How social verification makes the subjective objective. In R. M. Sorrentino & E. T. Higgins

- (Eds.), *Handbook of motivation and cognition: Foundations of social behavior* (pp. 28–84). New York: Guilford Press.
- Harkins, S. G., & Petty, R. E. (1987). Information utility and the multiple source effect. *Journal of Personality and Social Psychology*, 52, 260–268.
- 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.
- Insko, C. A., Thibaut, J. W., Moehle, D., Wilson, M., Diamond, W. D., Gilmore, R., et al. (1980). Social evolution and the emergence of leadership. *Journal of Personality and Social Psychology*, 39, 431–448.
- Jones, E. E. (1979). The rocky road from acts to dispositions. *American Psychologist*, 34, 107–117.
- 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.
- Kenworthy, J. B., & Miller, N. (2002). Attributional biases about the origins of attitudes: Externality, emotionality, and rationality. *Journal of Personality and Social Psychology*, 82, 693–707.
- 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.
- Kruglanski, A. W., Webster, D. M., & Klem, A. (1993). Motivated resistance and openness to persuasion in the presence or absence of prior information. *Journal of Personality and Social Psychology*, 65, 861–876.
- Maheswaran, D., & Chaiken, S. (1991). Promoting systematic processing in low-motivation settings: Effect of incongruent information on processing and judgment. *Journal of Personality and Social Psychology*, 61, 13–25.
- Matz, D. C., & Wood, W. (2005). Cognitive dissonance in groups: The consequences of disagreement. *Journal of Personality and Social Psychology*, 88, 22–37.
- 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.
- Milgram, S. (1974). *Obedience to authority*. New York: Harper & Row.
- Moskowitz, G. B. (1993). Individual differences in social categorization: The influence of personal need for structure on spontaneous trait inferences. *Journal of Personality and Social Psychology*, 65, 132–142.
- Neuberg, S. N., & Newsom, J. T. (1993). Personal need for structure: Individual differences in the desire for simple structure. *Journal of Personality and Social Psychology*, 65, 113–131.
- Nielsen, M. E., & Miller, C. E. (1997). The transmission of norms regarding group decision rules. *Personality and Social Psychology Bulletin*, 23, 516–525.
- O'Laughlin, M. J., & Malle, B. F. (2002). How people explain actions performed by groups and individuals. *Journal of Personality and Social Psychology*, 82, 33–48.
- Opp, K. D. (1982). The evolutionary emergence of norms. *British Journal of Social Psychology*, 21, 139–149.
- Prentice, D. A., & Miller, D. T. (1993). Pluralistic ignorance and alcohol use on campus: Some consequences of misperceiving the social norm. *Journal of Personality and Social Psychology*, 64, 243–256.
- 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.
- Reeder, G. D., Vonk, R., Ronk, M. J., Ham, J., & Lawrence, M. (2004). Dispositional attribution: Multiple inferences about motive-related traits. *Journal of Personality and Social Psychology*, 86, 530–544.
- Richter, L., & Kruglanski, A. W. (2003). Motivated closed mindedness and the emergence of culture. In M. Schaller & C. S. Crandall (Eds.), *The psychological foundations of culture* (pp. 101–122). Mahwah, NJ: Erlbaum.
- Schaller, M. (2001). Unintended influence: Social-evolutionary processes in the construction and change of culturally-shared beliefs. In J. P. Forgas & K. D. Williams (Eds.), *Social influence: Direct and indirect processes* (pp. 77–93). Philadelphia: Psychology Press.
- Schaller, M., Boyd, C., Yohannes, J., & O'Brien, M. (1995). The prejudiced personality revisited: Personal need for structure and formation of erroneous group stereotypes. *Journal of Personality and Social Psychology*, 68, 544–555.
- 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.
- Sechrist, G. B., & Stangor, C. (2001). Perceived consensus influences intergroup behavior and stereotype accessibility. *Journal of Personality and Social Psychology*, 80, 645–654.
- Shestowsky, D., Wegener, D. T., & Fabrigar, L. R. (1998). Need for cognition and interpersonal influence: Individual differences in impact on dyadic decisions. *Journal of Personality and Social Psychology*, 74, 1317–1328.
- Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. In S. Leinhard (Ed.), *Sociological methodology* (pp. 290–293). Washington, DC: American Sociological Association.
- Stangor, C., Sechrist, G. B., & Jost, J. T. (2001). Social influence and intergroup beliefs: The role of perceived social consensus. In J. P. Forgas & K. D. Williams (Eds.), *Social influence: Direct and indirect processes* (pp. 235–252). Philadelphia: Psychology Press.
- Thompson, M. M., Naccarato, M. E., Parker, K. C. H., & Moskowitz, G. B. (2001). The personal need for structure and personal fear of invalidity measures: Historical perspectives, current applications, and future directions. In G. B. Moskowitz (Ed.), *Cognitive social psychology: The Princeton symposium on the legacy and future of social cognition* (pp. 19–39). Mahwah, NJ: Erlbaum.
- Trope, Y., & Gaunt, R. (2000). Processing alternative explanations of behavior: Correction or integration? *Journal of Personality and Social Psychology*, 79, 344–354.
- Tweed, R. G., & Conway, L. G., III. (in press). Fundamental beliefs and coping across cultures. In P. T. P. Wong & L. C. J. Wong (Eds.), *Stress and coping: Multicultural & international perspectives*. New York: Kluwer Academic.
- Wegener, D. T., Petty, R. E., Smoak, N. D., & Fabrigar, L. R. (2004). Multiple routes to resisting attitude change. In E. S. Knowles & J. A. Linn (Eds.), *Resistance and persuasion* (pp. 13–38). Mahwah, NJ: Erlbaum.
- Wilder, D. (1977). Perception of groups, size of opposition, and social influence. *Journal of Experimental Social Psychology*, 13, 253–268.
- Wright, E. F., Luus, C. A. E., & Christie, S. D. (1990). Does group discussion facilitate the use of consensus information in making causal attributions? *Journal of Personality and Social Psychology*, 59, 261–269.
- Zuckerman, M., & Feldman, L. S. (1984). Action and occurrences in attribution theory. *Journal of Personality and Social Psychology*, 46, 541–550.

Received January 29, 2003

Revision received April 22, 2005

Accepted May 20, 2005 ■