

The False Enforcement of Unpopular Norms*

Robb Willer

University of California, Berkeley

Ko Kuwabara

Columbia University

&

Michael W. Macy

Cornell University

* Contact Robb Willer at Department of Sociology, University of California, Berkeley, CA 94720. E-Mail: Willer@Berkeley.edu. We would like to thank the members of Cornell's Laboratory Experimental Group for contributions to earlier versions of this paper, and the National Science Foundation for supporting this research, grant numbers SES0241657, 0602212, and 0405352.

THE FALSE ENFORCEMENT OF UNPOPULAR NORMS

ABSTRACT

Unlike laws, norms depend on informal enforcement by peers. Prevailing theory assumes that people enforce norms in order to pressure others to act in ways that they approve. Yet there are numerous examples of “unpopular norms” in which people compel each other to do things that they privately disapprove, including public support for oppressive regimes (Kuran 1995), infibulation (Mackie 1996), honor killings (Vandello and Cohen 2003), adoration of incomprehensible scholars (Willer 2004), and self-destructive adolescent behaviors (Prentice and Miller 1993). While peer sanctioning suggests a ready explanation for why people conform to unpopular norms, it is harder to understand why people would *enforce* a norm they privately oppose. We argue here that people enforce unpopular norms to show that they have complied out of genuine conviction and not because of social pressure. We use laboratory experiments to demonstrate this “false enforcement” in the context of a wine tasting (Study 1) and an academic text evaluation (Study 2). In both studies we found that participants who conformed to a norm due to social pressure then falsely enforced by publicly criticizing a lone deviant. Finally, we show that enforcement of a norm effectively signals the enforcer’s genuine support for the norm (Study 3). These results demonstrate the potential for a vicious cycle in which perceived pressures to conform and falsely enforce reinforce one another, a dynamic that can trap a population in a self-enforcing, unpopular norm.

INTRODUCTION

In November, 1978, the Reverend Jim Jones and over 900 of his followers living in a remote camp in Guyana committed mass suicide by drinking cyanide-laced Flavor Aid. On the brink of the “Jonestown Massacre,” a lone dissenter challenged Jones in front of the congregation about the merits of “revolutionary suicide” (Maaga 1998). A chorus of voices criticized Christine Miller for her dissent from the group’s normative judgment, insisting that she follow the decision of mass suicide. Miller lost her appeals, and almost everyone gathered would die within the hour. One report said that many of the prior opponents of the mass suicide were the first to line up to be poisoned. Miller herself relented and also ended her life.

The Jonestown Massacre has become the textbook example of mass persuasion, of how groups can be convinced of even the most extreme beliefs by a single charismatic leader (Cialdini 1988). However, Jones’ repeated use of suicide drills (“white nights”) suggests an alternative explanation that does not require the assumption that Jones’ followers acted out of ideological commitment to revolutionary suicide. “White nights” were mock suicide drills in which cult members did not know until after drinking the Flavor Aid that it had not been poisoned. Repeated drills may have led members to perceive drinking the Flavor Aid (and risking suicide) to be less costly than protesting and risking exposure as a deviant. Moreover, the hostile reaction to Christine Miller's dissent raises an equally provocative possibility – that skeptics might not only drink the poison but pressure others to do so as well. Why did Jones’ followers not only comply with the norm to drink the Flavor-Aid but also enforce it? Is it possible that they did so not out of conviction but instead to cover up their own private doubts and fears? Were

those who spoke up to criticize Christine Miller true believers, or were they looking for a way to prove their sincerity?

We will never know the answers to these questions about Jonestown, and many of the details remain hotly debated (Harray 1992). However, there are many other empirical cases of “naked emperors,” in which people feel pressured to publicly support behaviors or beliefs that they privately question. Our investigation focuses not on why people conform to unpopular norms in the face of social pressure. Rather, we want to know where the pressure comes from in the first place. Is it imagined? Does it reflect the conviction of true believers? Or is it possible that, as we suspect was the case in Jonestown, groups can become trapped in a self-enforcing equilibrium in which they pressure one another in order to cover up their own private doubts? If this “false enforcement” appears genuine to group members, it may appear so to researchers as well, leading to widespread underestimation of the number of unpopular norms.

Given the difficulty in natural settings of discerning genuine conviction from effective social posturing, we turn to controlled laboratory methods to carefully demonstrate false enforcement. We also investigate the effects of enforcement on perceptions of underlying conviction. Altogether, the results provide compelling support for the existence of false enforcement and its effectiveness in disguising private dissent.

The Normative Conception of Norms

Why do norms exist and why are they enforced? Despite deep differences in their theoretical approaches, functionalist, conflict, and utilitarian theories of social control

converge around the prevailing idea that norms arise because they are useful, either to society at large (in functionalist accounts), to dominant groups (in conflict approaches), or to those who enforce them (in utilitarian arguments). Parsons (1971:4-8) argues that norms “function primarily to integrate social systems,” a view echoed by Arrow who describes “norms of social behavior” as “reactions of society to compensate for market failures” (1971:22). Conflict theorists narrow the functionalist explanation by pointing to the role of norms in protecting the interests of dominant groups. In *The German Ideology*, Marx and Engels ([1845] 1986: 13) refer to norms as “ruling ideas” which “are nothing more than the ideal expression of the dominant material relationships, the dominant material relationships grasped as ideas; hence of the relationships which make the one class the ruling one, therefore, the ideas of their dominance.” The utilitarian conception is summarized by Hechter and Opp (2001: xvi): “The view that norms are created to prevent negative externalities, or to promote positive ones, is virtually canonical in the rational choice literature” (see also Bowles and Gintis 2001:6).¹

The utilitarian conception of norms is also canonical in research on social dilemmas – situations in which choices that are individually rational can lead to collectively irrational outcomes. Laboratory experiments have shown that norms obligating mutual cooperation in social dilemmas can emerge spontaneously in groups that would otherwise

¹ The utilitarian view does not require that norms be objectively useful so long as they are subjectively desirable, given some distribution of preferences. For example, working from a rational choice perspective, Horne defines norms as “rules, about which there is at least some degree of consensus, that are enforced through social sanctions” (2001a:5).

remain trapped in a deficient equilibrium (Yamagishi 1986; Fehr and Fischbacher, 2004). These studies suggest that norms provide solutions to collective action and coordination problems in social dilemma situations (Horne 2001a).

Increasingly, social psychologists, economists, and sociologists are questioning the canonical account and proposing an alternative possibility, one that invites us to revisit our sociological intuitions about informal social control. Norms do not necessarily solve social dilemmas or correct market failures, they can also undermine social welfare. “Unpopular norms” (Bicchieri and Fukui 1999) can emerge through a cascade of self-reinforcing social pressure that increases with the level of conformity (Centola, Willer, and Macy 2005). Like a witch hunt, the process can quickly spiral up into a powerful and dangerous dynamic.

The problem is highlighted in the Hans Christian Andersen fable of “The Emperor’s New Clothes,” which tells the tale of a naked emperor whose nonexistent clothes are widely admired by those fearful of being regarded as “unfit for office” ([1837] 1998). The “Emperor’s Dilemma” (Centola et al. 2005) formalizes the decision of whether to join a group consensus that violates private beliefs. Every player has the same four choices – whether or not to conform to an unpopular norm and whether or not to enforce it. The game qualifies as a social dilemma under the assumption that the payoffs to each player favor conforming and enforcing if all others do the same, yet each player would prefer that everyone deviates and tolerates deviance. However, the game is the opposite of most other social dilemmas, describing situations in which norms of cooperation are not useful for the group, but instead are harmful.

Below we present two prominent explanations of how a population can become trapped in an unpopular norm: *herd behavior* and *pluralistic ignorance*. We then consider an alternative explanation based on *false enforcement* that better explains the remarkable stability of some unpopular norms. Finally, we report results of three controlled experiments designed to investigate the existence and dynamics of false enforcement.

EXPLAINING UNPOPULAR NORMS

Explanations of unpopular norms can be classified into three types, based on three different causal mechanisms – *herd behavior* (primarily studied by economists), *pluralistic ignorance* (primarily studied by psychologists), and *false enforcement* (proposed by Centola et al. 2005). Herd behavior (Banerjee 1992) and information cascades (Bikhchandani, Hirshleifer, and Welch 1992, 1998; Hirshleifer 1995) are a response to widespread uncertainty that leads people to follow the decisions of others who are assumed to have more accurate or reliable information, when in fact, the others are also responding to uncertainty by following the herd. The belief that “this many people can’t be wrong” can generate a “Matthew effect” (Merton 1968), as when people become more likely to download an unfamiliar song the more they see that others download it (Salganik, Dodds, and Watts 2006; Salganik and Watts 2008; Hanson and Putler 1996). Accounts based on herd behavior have also been used to explain instabilities in financial markets, such as bubbles and crashes (Lee 1998), bank runs (Chen 1999), and faddish investments (Schiller 2000; Graham 1999).

Herd behavior is highly dependent on the behavior of early movers who exert a disproportionate influence on the future direction of the collective (Arthur 1990). As a consequence, the behavior of the group tends to reflect the preferences of a small

minority of the population. Thus, by “following the crowd,” people are led to decisions that are sometimes inconsistent or even contrary to what they privately believe.

Pluralistic ignorance describes the “belief that one’s private attitudes and judgments are different from those of others, even though one’s public behavior is identical” and is common in situations where public behavior is widely misrepresented (Miller and McFarland 1991: 287). Like herd behavior, pluralistic ignorance (Katz and Allport 1931; Miller and McFarland 1991; O’Gorman 1986) is based on a false belief that becomes self-reinforcing. The difference between the two phenomena centers on Deutsch and Gerard’s (1955) distinction between “normative” and “informational” social influence. Informational influence rests on the need for accuracy, while normative influence reflects the need for social approval. With herd behavior, people address their uncertainty by copying the behavior of others who are assumed (incorrectly) to have better information. With pluralistic ignorance, people suppress their dissent and copy the behavior of others who are assumed (again incorrectly) to represent the popular majority. It is, in Krech and Cruchfield’s (1948: 388-9) words, the situation where “no one believes, but everyone thinks that everyone believes.” Andersen’s fable is a classic illustration: privately everyone knows the emperor is naked, but everyone goes along with the public charade, based on the false belief that others are sincere in their public expressions of fawning adoration.

The appeal of Andersen’s fable reflects the ubiquity of “naked emperors.” For example, studies show that college students believe they are expected to celebrate intoxication and participate in binge drinking rituals as a test of group loyalty (Nagoshi et al. 1994; Perkins and Wechsler 1996; Baer 1994; for a review, see Borsari and Carey

2001). Yet Prentice and Miller (1993) found that most students were privately less comfortable with excessive drinking than they believed others to be. The widespread misperception that one is in a minority in turn leads to a “spiral of silence” (Noelle-Neumann 1984) that confirms the belief.

Explanations based on herd behavior and pluralistic ignorance posit self-reinforcing beliefs that are quite fragile, as in the Andersen fable. In both theories, the unpopular norm is likely to crumble if more accurate information on the underlying consensus comes to light. The norm is a house of cards that can quickly collapse if people discover that the belief that motivated their conformity was false – e.g., that the child who laughs at the Emperor is admired and not ridiculed.

This fragility may not be the case, however, if the norm is enforced. The threat of informal sanctions by peers can allow unpopular norms to become surprisingly stable, as the following everyday examples suggest:

- *Homophobia*: Derision of homosexuals by men seeking to affirm their gender identity in the face of masculine insecurity (Willer 2005).
- *Snobbery*: Gossiping about a *faux pas* or ridiculing those who fail to appreciate esoteric artwork, by intellectual and cultural snobs anxious to affirm their superior intelligence, breeding, and aesthetic taste.
- *Adolescent rebellion*: Students chiding one another for buying into school achievement instead of engaging in self-destructive behaviors (e.g., shoplifting, reckless driving, binge drinking).

- *Witch trials*: Affirming the credibility of one's confession by revealing the names of other religious or political heretics, thereby perpetuating the anxiety that fuels the accusations.
- *Unintelligible scholarship*: Joining the chorus of praise for an obtuse scholar to address one's own insecurity fostered by not understanding an unintelligible text that others appear to admire.
- *Flag waving*. Liberal Democrats voting for the war in Iraq, despite deep private misgivings, to avoid appearing unpatriotic and to prove their "fortitude," thereby increasing the vulnerability of anyone tempted to dissent.

These examples of "naked emperors" are highly stable because they rest on false beliefs supported by very real willingness to enforce. What is most puzzling, however, is the possibility that that these norms are publicly enforced by people who privately question the norm. We know why people enforce norms they agree with – to get people to behave the way they want them to. But why would people publicly enforce a norm that they secretly wish would go away?

THE ILLUSION OF SINCERITY

"Naturally, the best proof of the sincerity of your confession was your naming others whom you had seen in the Devil's company."

-- Arthur Miller, 1996

Centola, Willer, and Macy (2005) propose a simple explanation: Enforcement is an effective way for those who conformed to demonstrate their sincerity – i.e. to show that they did not conform simply to secure social approval. This is because, to true believers

in the group, mere conformity is insufficient. Those who conform may only be posturing for group approval and can only be trusted to do so as long as behavior is monitored (Hechter 1987). The demand for sincerity was fundamental to the anti-Communist “Red Scare” of the 1950’s. Loyalty oaths for government employees would often require that the oath-taker do so sincerely. For example, University of California employees are still required to sign a form certifying not only that they will “support and defend the Constitution of the United States...against all enemies,” but also that they “take this obligation freely, without any mental reservation,” and not simply because it is a precondition for employment. Sincere conformity was also demanded by the Catholic Church during the Spanish Inquisition. It was required that confessions given following torture be sincere, denoted by the phrase “*confessionem esse veram, non factam vi tormentorum*” (i.e. the confession was true and free).

The problem for University of California job applicants, heretics seeking to avoid torture, or more commonly, adolescents afraid of being exposed as “posers,” is that sincerity is difficult to signal. Even sincere conformists might worry that others may suspect they did not “take this obligation freely.” Research on the “illusion of transparency” underscores the everyday anxieties motivating individuals to create an impression of sincere conformity (Gilovich et. al. 1998). The illusion of transparency refers to the tendency of individuals to overestimate others’ abilities to view their internal states. Of particular significance, this research demonstrates that liars overestimate how detectable their lies are to others. This research implies that individuals who conform out of a desire for social approval rather than genuine conviction will worry that others may detect their insincerity.

One way out of this predicament is to demonstrate sincerity by behaving the way true believers do, and sanction deviants in the group who voice dissent from the norm. Centola et al. (2005) call this signaling behavior the “illusion of sincerity,” which they illustrate with two examples, college drinking and scholarly posturing:

1. Insecure students who worry about social acceptance may seek peer approval by drinking heavily and celebrating intoxication. However, they must be careful not to appear to be motivated by this goal or they risk being scorned as a “poser.” Thus, it is not enough to drink and party – one must also target social approval toward those who drink and disapproval at those who do not, which in turn adds to the peer pressure that leads others seeking social acceptance to join in.

2. Clance’s (1985) “imposter phenomenon” refers to the anxiety among many academics (and other professionals) that one is an “intellectual fraud.” Those who affirm their erudition by pretending to appreciate some highly opaque scholar (Gross and Levitt 1994; Sokal and Bricmont 1998) worry that others will see through the bluster. A common solution is to disparage as intellectually shallow those unable to appreciate real genius. This in turn adds yet another voice to the chorus of intimidation that produces the insecurity motivating the behavior in the first place.

In both these examples, the norm becomes self-enforcing. Most norms are self-limiting, which means the motivation to enforce decreases with the rate of conformity. That is because the motivation to enforce derives from what Heckathorn (1988) calls a “regulatory interest” in changing the behavior of others. If the marginal return on each

additional act of enforcement diminishes with the rate of compliance, then the more who comply, the lower the incentive to enforce.

Unpopular norms are unusual in that they are enforced by those with no true regulatory interest, which is why some functionalists and utilitarians may have difficulty believing such norms can exist. Axelrod (1986) proposed “metanorms” as an explanation for norm enforcement when regulatory interests are weak, that is, when the cost of administering sanctions exceeds the benefit of sanctioning, creating “second-order” free-riding (Oliver 1980). A metanorm is a norm that obligates enforcement of another norm. A familiar example is the metanorm against tolerance of racial prejudice, or honor systems that obligate group members to report violations. Recent laboratory studies have found compelling evidence for metanorms under controlled conditions.²

People may also enforce norms they secretly oppose in order to create an “illusion of sincerity” that they support the norm. Just as conformity may be regarded as insincere, so too may enforcement be seen as insincere or externally motivated if there is an explicit obligation or pressure to enforce, as in an honor system. However, in the absence of such

² For example, in a series of experiments with public goods games, Yamagishi (1986, 1988a, 1988b) demonstrates that sanctioning systems are established in direct response to the likelihood of free-riding. Horne (2001b, 2004; 2009) shows that regulatory interests can motivate both sanctioning and sanctioning of sanctioning. Fehr and colleagues (Fehr and Gächter 2000, 2002; Fehr and Fischbacher 2004) go further, suggesting that metanorms are also rooted in strong emotional reactions that drive cooperators as well as mere observers to punish non-cooperators, even at a net cost and without promise of long-term gain. Similarly, Ostrom, Walker, and Gardner (1992) note that, in the absence of in-group communication, participants in public goods games can overuse sanctions. Shinada, Yamagishi, and Ohmura (2004) extend the idea of “altruistic punishment” to an evolutionary theory of group cohesion, showing that non-cooperation by in-group members is more likely than non-cooperation by out-group members to trigger costly sanctions.

obligations, “voluntary” enforcement (in the way that “true believers” are expected to behave) may provide an effective signal that one’s behavior is motivated by genuine conviction.

Centola et al. (2005) used agent-based modeling to examine the population-level implications of false enforcement as a signal of sincerity. In their model, a very small fraction of true believers can spark a cascade of conformity and false enforcement that quickly engulfs a vulnerable population – not because people are converted to new beliefs, but just the opposite – because of the need to affirm sincerity felt by those who know that their conformity is a lie.

EMPIRICAL OVERVIEW

While computational experiments demonstrated how the desire to create an “illusion of sincerity” can generate cascades of false enforcement of unpopular norms, the behavioral assumptions of the model have not been tested under controlled conditions. The experimental studies that follow extend this work by testing two key behavioral assumptions – that conformity to an unpopular norm can lead in turn to false enforcement, even without an explicit metanorm (Studies 1 and 2), and that false enforcement can effectively signal that one is a “true believer” in the norm (Study 3). More specifically, in Study 1 we demonstrate false enforcement in a wine-tasting setting in which participants must decide whether to agree with a consensus evaluation that participants had reason to doubt, followed by a judgment of the performance of fellow participants. Study 2 establishes the robustness of false enforcement in an academic text evaluation setting, under conditions that differed in several ways from Study 1. Finally, Study 3 investigates whether enforcers of norms are seen as “true believers” who

conform because of underlying conviction, and not simply to gain approval. Taken together, the studies illuminate the potential for a vicious circle of false admiration for naked emperors and false derision of those who cannot see the clothes.

STUDY 1: FROM CONFORMITY TO FALSE ENFORCEMENT

“Fanaticism is overcompensation for doubt.”

-- Robertson Davies

Study 1 tests the claim that a group consensus can lead participants to not only conform to a norm they privately do not support, but also enforce compliance by others. To do so we use a variant of the classic Asch (1951) conformity experiment. In Asch’s study, participants were asked to identify which of three lines was the same length as a focal line. One of the three lines was identical and the other two were obviously shorter or longer. On several “critical trials” participants were asked to respond after confederates had all given an identical but blatantly incorrect judgment. 32% of participants’ answers on critical trials were incorrect, and 74% of participants gave at least one incorrect answer. When participants were asked to compare the lines in isolation, the error rate did not depart significantly from zero, indicating that incorrect answers were a consequence of group influence.

Study 1 uses a new setting for studying not just conformity to, but also enforcement of social norms. In our experimental setting, participants were asked to taste and evaluate three wines and then evaluate the performance of five fellow wine tasters via a computerized interface. Although participants were told that they were tasting three

different wines, in fact all three samples were actually from the same bottle. The only difference among the three samples was that one wine was secretly tainted with vinegar. In part one of the experiment, participants rated the wines after reading the erroneous responses of four computer-simulated confederates and before reading the correct response of a simulated deviant. This setting allowed us to establish a normative rating of the wines, and then create a lone deviant in order to study participants' enforcement behavior. The norm was that one of the two identical wines was greatly inferior to the other and even worse than the wine tainted with vinegar. After tasting and rating the wines, participants were then asked to judge the wine tasting abilities of all the group members, including the simulated deviant who submitted ratings after the participant. All wine ratings were public, but judgments of other participants were public for half the participants and private for the other half. Participants were randomly assigned to one condition or the other.

We define false enforcement as the public enforcement of a norm that is not privately endorsed. We have argued that false enforcement can occur when individuals have conformed to the norm due to social pressure and are motivated to demonstrate the sincerity of that conformity. Thus, not all participants in the experiment were expected to falsely enforce – only those who conformed to the norm by agreeing with the judgment of their peers.

Study 1 was designed to identify false enforcement by comparing participants' peer judgments in the public and private conditions. In behavioral research, directly asking participants to reveal their beliefs creates a perceived pressure from the researcher to

behave consistently with those reported beliefs (e.g., Tedeschi, Schlenker, and Bonoma 1971; Gaes, Kalle, and Tedeschi 1978). We therefore avoided a direct measure of private belief.³ Instead, we had some participants express their peer judgments anonymously, where these judgments should reflect their private beliefs about their peers, while others expressed their judgments in public, where judgment could be influenced by perceived social pressure to demonstrate the sincerity of conformity.

When enforcement is anonymous, we predict that peer judgments will reflect participants' private beliefs about the wines – either their agreement or disagreement with their peer's wine ratings. Other factors might also influence peer judgments, such as admiration of the courage to challenge the group consensus, as suggested by research on minority influence (Moscovici and Nemeth 1974) or the need for cognitive consistency (Festinger 1954) between public wine ratings and private peer ratings. However, these effects should be constant across all conditions and therefore should not affect the public-private comparison.

Further, we expect the peer judgments of those who did not conform to reflect agreement with the lone deviant, and disagreement with the conformist wine tasters, in both the public and private sanctioning conditions. Having publicly disagreed with the

³ Future research could try different methodological strategies for assessing private beliefs, perhaps using non-obtrusive measures like the “bogus pipeline” (Jones and Sigall 1971; but see also Aguinis and Handelsman 2006).

normative consensus, participants who did not conform need not prove that their evaluations of the wines were genuine.

Among participants who conformed to the norm, however, we predict a difference in public and private peer judgments. When given anonymously, their peer judgments can be expected to reflect their private disagreement with the conformists about the quality of the wines. When given publicly, their peer judgments provide an opportunity to show the other group members that their wine ratings reflected genuine conviction and not simply “going along” to avoid a negative social judgment from one’s peers. Hence, those participants who conformed to a norm they privately did not agree with are predicted to favor the conformists over the deviant in public, but not in private. If we find that conformists favor the deviant in private but not in public, while non-conformists favor the deviant in both private and public, it would provide important support for the theory of false enforcement.

We tested our prediction using a repeated-measures ANOVA of participant’s peer judgments, using three predictors: how much the participant conformed, whether the peer being judged conformed or not, and whether peer judgments were given in public or private. We predict a 3-way interaction in which study participants’ conformity will tend to lead to more favorable judgments of conformist peers (relative to the deviant peer), but only when the judgment is given in public. If this interaction term is not significant, the false enforcement hypothesis must be rejected.

METHOD

Fifty-two undergraduate students (34 women, 18 men)⁴ at a large private university participated in the study in return for \$8 or extra credit in a sociology class. Participants were recruited either by fliers advertising payment for participation in a sociology study or via an announced extra credit opportunity in their sociology class. Participants were directed to an internet-based sign-up page where they were first asked to take a brief “Art Appreciation Test” in which they indicated which of several pairs of abstract paintings had been judged by experts as the “100 best paintings of the 20th century.” Regardless of participants’ responses, all were told that their score qualified them for participation in a “High Art Appreciation Group.” The test was in fact bogus and designed to heighten anxiety about social judgments based on sophisticated cultural tastes.

Participants arrived at the lab and participated in groups of six at a time.⁵ A research assistant seated each participant at a separate computer terminal in a computer classroom with cardboard dividers arranged to prevent eye contact between participants. Though participants believed that they were interacting with the other 5 participants in the room, in fact all interacted with the same computer-simulated “confederates” for the duration of the study. Participants were told they would receive all instructions on their computer terminals. Once everyone was seated the study session began. Instructions displayed on

⁴ One participant’s data was lost due to a computer crash. Neither gender nor age ($M = 20.1$ years) of participants significantly interacted with any of the results presented and are not discussed further.

⁵ Two study sessions included only five participants due to no-shows. In both cases the sixth position was filled by an undergraduate research assistant posing as a participant.

their terminal indicated that they were to participate in an “Aesthetic Discrimination and Appreciation Study” funded by the National Endowment for the Arts. They were told that the study concerned the relationship between aesthetic tastes in different domains (e.g. ability to appreciate fine art and wine) and that in the present study the researchers were particularly interested in their opinions of several non-alcoholic wines.

Next to their computer screen, each participant found a set of three cups of non-alcoholic wine marked A, B, and C.⁶ Their task was to rate these three wines. To enhance concerns about others’ social judgments of their performance, participants were told (1) that all group members would participate in a face-to-face group discussion after the computerized component of the study, and (2) that their scores in the study would be evaluated according to the “Beckman Aesthetic Discrimination Scale” and score reports on all study participants would be mailed to any participants interested in receiving one. In fact there is no such thing as a “Beckman scale” and there was no post-study discussion or score report.

Participants were further instructed that all six participants would partake in a group wine-tasting procedure known as the “Dutch Round” that had been adapted for a computer-mediated environment. Participants were told that the Dutch Round was a centuries-old tradition among Dutch wine enthusiasts in which people stood in a circle and took turns giving their opinions of different wines, followed by a second round in

⁶ University and state regulations prohibited the use of alcoholic wines with undergraduate participants. We tested a large number of non-alcoholic wines before finding one that unambiguously tasted better without the added vinegar. It is possible that non-alcoholic wine invites less snobbish behavior than would a “real” wine and therefore provides a more conservative test of the hypothesis, since snobbery can be expected to induce increased anxiety about negative social judgments.

which they evaluated one another's wine tasting ability, again in turn. In fact no such Dutch tradition exists and the details of the Dutch Round were invented to create the conditions necessary for our experiment.

Participants were then given an opportunity to "introduce" themselves to one another by answering questions about their gender, year in school, college affiliation, what musical instrument they played (if any), what foreign language they spoke (if any), and whether they had taken a popular wine class offered at the university where the study was conducted.

Following the introductions, participants were told that they had been randomly assigned the ID number 5, which meant that their personal information was displayed 5th out of the 6 participants. Responses from all six people were displayed on the screen, in the order to which each had been assigned, but in fact five of the six responses, other than the participant's, were simulated. The other five "participants" were all introduced as upperclassmen with proficiency with a musical instrument and foreign language. Additionally, three of the five simulated confederates were identified as having taken the wine class, with the exception of one of the first four confederates and the confederate in the 6th position, who would become the lone deviant.⁷ This information was provided to heighten the participant's concerns about social judgments from those with what appeared to be superior cultural credentials. The wine class was used to add credibility to the group consensus.

⁷ Participants also indicated whether they had taken the wines class. Taking the class did not affect any of the results presented here.

At this point in the program, participants were told that all group members would taste three different wines in the first part of the study and then submit ratings of them in order of their previously assigned ID numbers. Each of the three samples contained 2 oz. of non-alcoholic wine. Instructions prompted participants to taste all three wines twice each, in accordance with the tradition of the Dutch Round. Participants were then asked to rate the three wines on the following dimensions: bouquet, flavor, aftertaste, robustness, and overall. The rating system was the familiar five-point A to F scale used in the classroom.

All participants were assigned to the fifth position in the Dutch Round, which meant that they submitted their wine ratings after seeing four others' ratings,⁸ and before the last person's ratings were displayed. Table 1 shows the ratings given to each of the three wines by the 4 confederates and the simulated deviant, in the order that they took turns. As summarized in the table, all four previous raters (simulated on the computer) gave very high ratings to Wine A, and very low ratings to Wine B, which was in fact identical. Indeed, they gave somewhat higher ratings to Wine C, which had been tainted with vinegar. The last confederate, whose turn came after the participant, broke ranks with the consensus and rated Wine B the same as Wine A.

[Table 1 about here]

In the Asch study, confederates occasionally gave correct answers to ensure credibility of the experimental design; we used Wines A and C for the same purpose. All

⁸ Asch found at least three unanimous confederates were needed to induce participant conformity in his setting (1951). Unlike Asch, we ran the experiment on computers, using simulated participants as confederates. We use the term "confederate" to distinguish their behaviors from those of the participant.

five confederates agreed on these two wines, including the lone deviant. However, they disagreed sharply about Wine B, with the deviant challenging the consensus view that Wine B was much worse than Wine A when in fact it was identical to that wine. As in the Asch experiment, we tested the stimuli using a separate set of individuals, outside of the group setting. This test ($N = 17$) confirmed that, in the absence of social pressure, ratings of the three wines were most similar to those given by the lone deviant. Though there was no significant difference in ratings given to Wines A and B ($p > .65$), both were rated significantly higher than Wine C ($p < .001$).

Following Part 1, participants proceeded to Part 2 of the Dutch Round, in which they were told that everyone would rate each others' wine rating abilities on five point scales ranging from A to F. The study featured a two-condition (private or public sanctioning) between-subject design. Half the participants were randomly assigned to the "private sanction" condition, in which they were told that their ratings of the other five wine tasters would be completely anonymous and private. The other half of participants were randomly assigned to the "public sanction" condition. These participants were told that ratings of the other five wine tasters would be submitted publicly in random order. The participant was then "randomly" chosen to go first in submitting his/her ratings of the others' wine-rating abilities. This was necessary to insure that enforcement behavior was not influenced by having observed enforcement behavior by the confederates, and to allow the participant to use enforcement to gain approval from others.

After submitting their performance ratings, participants were shown the performance ratings that the other five wine tasters had ostensibly submitted. These were simulated to be consistent with their wine ratings in Part 1, with four of the five

confederates giving negative judgments of the lone dissenter. None of the confederates offered negative judgments of the participant.

POST-STUDY QUESTIONNAIRE AND DEBRIEFING

In both conditions, after submitting their ratings of the others, participants completed a post-study questionnaire. Participants were individually debriefed about the actual goals of the research. Participants were asked if they had heard about the experiment from former participants. A funnel debriefing procedure was used to assess suspicion of deception in the study.⁹ The reasons for not revealing the true nature of the study were carefully explained in detail. Participants were further assured that their data would be entirely anonymous and that they could request a copy of the research findings.

Participants were asked to not discuss the details of the study with other students until the conclusion of the semester.

RESULTS

Over fifty years of research on conformity have demonstrated that participants can be influenced by group members to publicly misrepresent their private beliefs (Bond and Smith 1996). Although not our primary goal, we first tested this assumption since we needed participants to conform in order to see whether the same social pressures inducing

⁹ The level of suspicion was not high and did not significantly affect the results presented below.

conformity might also lead to false enforcement. All analyses of participant's wine ratings used participants' ratings in the "overall" category.¹⁰

The consensus among the four confederates who rated the wines prior to the participant created sufficient social pressure to generate conformity. Although Wines A and B were in fact identical, Wine B ($M = 3.96$) was rated significantly worse than Wine A ($M = 4.39$) ($t = 4.2, p = .004$).¹¹ Though Wine B's average rating was much higher than the consensus opinion of Wine C (2.25), this is because participants ranged from strongly conformist to strongly non-conformist. Twenty-seven participants (52.9%) rated wine B below the identical Wine A. Of these, three participants (5.9%) also rated Wine B even worse than the vinegar-tainted Wine C. Of the other participants, thirteen (25.5%) rated Wines B and A equally, while eleven rated Wine B higher than Wine A (21.6%).

Our main theoretical interest is not conformity but enforcement behavior. Can social pressure to conform lead participants not only to join the group consensus but also to enforce a view that they privately do not believe? To test for this, we submitted participants' peer judgments of the other wine tasters to a repeated-measures ANOVA. The model included two between-subject variables: the experimental manipulation (public or private enforcement) and the participant's conformity. Conformity was measured by the rating assigned to Wine B, for which the consensus was in error. Since

¹⁰ We repeated all analyses that follow using composite measures of participants' ratings of each wine, averaged across each of the different rating categories (e.g. bouquet, robustness, etc.). All results were substantively identical.

¹¹ Participants evaluated the wines before knowing whether their evaluations of the others would be public or private (which they were told just prior to evaluating the other wine tasters). Hence, rates of conformity should not have varied by condition, and it did not.

Wine B was identical to Wine A, we used Wine A as the baseline and measured conformity as the rating of Wine A minus the rating of Wine B.¹² A third predictor was a within-subjects designation of the target being judged (conformist or deviant). Since each participant rated five targets, the repeated measures dataset consists of 255 observations. Our theory of false enforcement does not predict any main effects. Rather, we predict an interaction effect in which greater conformity leads participants' to give lower ratings of the lone deviant relative to the conformist targets, but only when peer judgments are public.

The results of the ANOVA are summarized in Table 2. As shown, we found two significant effects. The two-way interaction of deviant target and participant's conformity indicates that the more strongly participants conformed, the more they tended to favor conformist targets $F(1, 196) = 2.85, p < .04$. The three-way interaction of deviant target, participant's conformity, and experimental condition represents the additional effect of peer judgments being submitted in public instead of private, $F(3, 196) = 3.16, p < .03$. This interaction effect indicates that the effect of participant's conformity on favoring conformists over the lone deviant was significantly stronger in the public condition.

[Table 2 about here]

These interaction effects are evident in Figure 1, which reports enforcement behavior broken down by whether the peer judgment was given in public or in private,

¹² Though participants' ratings of Wine C may also have been influenced by the first four confederates' wine ratings, no confederates defected from this normative rating and thus it is not relevant to the possibility of subsequent false enforcement.

whether the target was conformist or deviant, and whether the participant had previously conformed. Although the ANOVA used a continuous measure of conformity, for purposes of illustration, we classified participants as “conformists” if they agreed with the consensus that Wine A was superior to Wine B and “non-conformists” otherwise.

[Figure 1 about here]

Figure 1 shows that participants favored the conformist targets over the deviant only when the participant had conformed and only when their peer judgments were given in public. In private, conformist participants favored the deviant, as did the non-conformist participants. These results provide clear support for the prediction that conformists will give unfavorable peer judgments to deviants, compared to conformist targets, only when evaluations are made in public. In private, they favored the deviant, as did those participants who refused to conform to the group consensus.

DISCUSSION OF STUDY 1

Past research has shown that social influence can cause people to agree with a group consensus that they privately know to be false (Asch 1951). However, much less is known about how the pressure to agree is generated. We wanted to know if a group consensus could cause people not only to publicly express views they privately questioned, but also pressure others to join in the charade.

If participants were like the characters in the Andersen fable, they would have been praised the lone deviant, who plays the role of the child whose laughter breaks the spell. Indeed, they might penalize the four confederates. However, that is not what we observed. Consistent with our theory of false enforcement, we found that participants not

only complied with the normative opinion to favor one wine over another when they were actually identical, but also publicly sanctioned a lone deviant in the group for dissenting from the norm. The ANOVA showed that conformity led participants to publicly favor conformist peers over the lone deviant. However, when enforcement was private, those who had conformed at higher levels approved of the deviant whose views corresponded to their private beliefs. These results support the hypothesis that social pressure to conform to a norm can lead those who conform to also sanction others for not conforming. The implication is that unpopular norms can become self-enforcing, even without formal institutional arrangements, such as an honor system with explicit metanorms mandating enforcement. This in turn allows cascades to trap a population into worshipping a naked emperor, without the option for a child to break the spell.

One might wonder whether participants in the study who judged the conformist targets as more competent wine raters than the deviant target might have privately agreed with the conformist targets they praised; i.e., perhaps these peer judgments were sincere and not the product of social pressure. Results, however, are inconsistent with that explanation. First, Wines A and B were in fact identical, and pretesting of these stimuli confirmed that in the absence of social pressure, participants on average rated them as about equal quality. Yet the consensus of the first four confederates was not just that the wines were unequal, it was that the two wines were at opposite ends of the scale, with Wine B rated on average even lower than a wine tainted with vinegar. None of the pre-testers came close to such a view. Thus, it is unlikely that participants who rated the four confederates as more competent wine tasters did so because they sincerely believed that Wine A was vastly superior to Wine B or that Wine B tasted as bad or worse than a wine

tainted with vinegar. Second, we found that participants randomly assigned to give peer judgments in private tended to rate the deviant more positively than the conformist targets, indicating that, in the absence of social pressure, participants would not have agreed with or enforced the consensus view.

Still another possibility is that participants were highly uncertain regarding the stimulus, so much so that they were heavily influenced by the majority of wine-tasters and came to accept the consensus wine rating. This explanation, however, also fails to account for the observed difference in public versus private peer judgments. If participants were in fact influenced by the majority view, their peer judgments should have reflected their revised beliefs both in public and in private.

Nonetheless, results of a pilot version of the experiment ($N = 18$) suggest the possibility that uncertainty about the norm may be an important factor in false enforcement. In the pilot study, the tainted Wine C was given high marks by the four confederates while the untainted Wine B was given very low marks. However, the superior rating of Wine C was so at odds with reality that participants rejected this norm, rating Wine B ($M = 3.83$) significantly higher than the vinegar-tainted Wine C ($M = 2.78$) ($t = 2.19, p < .05$). This pattern mirrors the assessments given by anonymous raters in stimulus testing. In other words, the consensus that Wine C was a superior wine had no measurable effect on the participants. The explanation for this may be the absence of uncertainty in the pilot study. In stimulus testing, raters almost unanimously rated Wines A and B above Wine C, but results were far more mixed on the relative quality of Wine B, which some rated higher than Wine A and some lower. The greater uncertainty about

the quality of Wine B may have lowered participants' commitments to their private beliefs, freeing them up to yield to social pressure.

This explanation, however tentative, is consistent with similar findings on the effects of uncertainty in promoting conformity (Deutsch and Gerard 1955). Had the consequences of deviance been more severe, it is possible that conformity might have been higher in the pilot study, but with relatively weak social pressure, high levels of certainty that a norm is false appears to prevent conformity and false enforcement. Some uncertainty about the norm may therefore be a necessary condition for false enforcement in the absence of more intense pressure to conform.

Study 2: ENFORCING APPROVAL OF UNINTELLIGIBLE TEXTS

Study 1 demonstrated that social pressure can induce false enforcement of group norms. Participants not only complied with a consensus evaluation to rate one wine far lower than another when they were actually identical, but also publicly sanctioned a lone deviant in the group. Study 2 was designed to demonstrate the robustness of false enforcement under different conditions. In place of a wine-tasting, participants were asked to read an unintelligible academic text, ostensibly written by a respected academic. Using a similar cover story to that of Study 1, four confederates in succession gave positive evaluations of the text prior to the participant's turn. After the participant evaluated the text, a lone dissenter from the norm submitted a very negative evaluation of the text. In a second stage of the study, participants submitted judgments of the other

group members. Depending on experimental condition, this peer judgment was either given in public or in private. As in Study 1, we predict that participants who conform to the normative evaluation of the text will also enforce it, but only in the public condition.

This setting differs from that of Study 1 in several ways. First, in Study 1, participants were asked to rate three wines, and they were pressured to indicate that one of two identical wines was worse than a third wine that was tainted with vinegar. In Study 2 participants were pressured to say that a nonsensical text had high intellectual merit, an evaluation that is absolute rather than comparative. Second, in Study 1, evaluations were largely aesthetic, while in Study 2 participants were asked to give an analytical evaluation of an academic text. Finally, Study 1 used a pre-test to confirm that participants would rate the two wines equally when giving their ratings anonymously, while Study 2 included an additional anonymous text-evaluation condition in which participants evaluated the text privately. The anonymous text-evaluation was used to test our assumption that, in the absence of social pressure, participants would correctly recognize that the text was incomprehensible.

METHODS

Seventy-six undergraduate students (57 women, 19 men) at a large, public university participated for extra credit in a sociology class. Participants were recruited via an in-class announcement and scheduled for a specific study session by phone. Participants were scheduled in groups of six, this time to participate in a “Roundtable Text Evaluation

Study.”¹³ Participants were told that they were participating in a large-scale, multi-site study of peer review practices in academia involving undergraduates, graduate students, and faculty from “top schools.” Participants were informed that they would participate in a group text evaluation with five other undergraduates from their school in a computerized version of the “Bloomsbury Literary Roundtable.” The roundtable provided a credible cover story for the structure of the text evaluation. Similar to Study 1, participants were told that this was a traditional format wherein “rounders” would evaluate an academic text in order, and then take turns rating one another’s evaluation skills. In fact, no such text evaluation tradition exists.

As in Study 1, participants were told that they would be rated via the “Beckman Textual Discrimination Scale” and that ratings for all study participants would be published in a report available to all other study participants. Next, participants were given an opportunity to “introduce” themselves to the other participants over the computer network, though again all others were computer simulated. Participants were asked to indicate their age, year in school, approximate GPA, whether they intend to go to graduate school, and whether or not they had completed any philosophy classes. Participant 6, the deviant, was characterized as a Freshman with a GPA in the 2.75-2.99 range, no exposure to philosophy courses, and no intention to pursue a graduate degree, while Participants 1 through 4 were somewhat older, boasted higher GPAs, had taken philosophy, or intended to pursue a graduate degree. This information was provided to

13 One study session included only four participants due to insufficient sign-ups and a no-show. In this case the fifth and sixth positions were filled by undergraduate research assistants posing as participants.

increase the credibility of the normative consensus and heighten participants' concerns about social judgments from those with apparently superior academic credentials.

Next, participants were presented with a short essay to read and evaluate, ostensibly authored by “Dr. Robert Nelson, Ph. D., Albert W. Newcombe Professor of Philosophy, MacArthur ‘Genius Grant’ Recipient, 2001, Harvard University.” However, the essay was in fact an excerpt from “Differential Topology and Homology” and “Manifold Theory: (W)holes and Boundaries,” sections of Alan Sokal’s “Transgressing the Boundaries: Towards a Transformative Hermeneutics of Quantum Gravity” (1996). The text was written by Sokal to be unintelligible, intended to expose academic journals that would accept an unintelligible text for publication. Previous research has shown that participants rated this text as higher quality and more intelligible than they would otherwise when they believed it had been authored by a renowned Harvard professor (Willer 2004). Given the unintelligibility of the Sokol text, the distinguished authorship was used to lend credibility to the positive evaluations given by the simulated confederates.

After reading the essay, participants were asked to evaluate the text in several categories: “understandability,” “persuasiveness,” “logical coherence,” “quality of argument,” and “overall.” In one condition of the study, the “private evaluation” condition, participants were told that their text evaluations were private. In this condition, participants entered their evaluations and then proceeded to the post-study questionnaire. For the other two conditions, participants gave their text ratings fifth out of the six participants. As shown in Table 3, the first four simulated confederates all uniformly

praised the text, awarding As and Bs, but the sixth rater (the “lone deviant”) gave it very low ratings (Cs and Ds).¹⁴ In both of the public text evaluation conditions, participants then proceeded to the peer judgment task. In the “public evaluation/private sanction” condition, participants submitted their peer judgments privately. In the “public evaluation/public sanction condition” participants gave their peer judgments publicly and were told that they were randomly chosen to go first of the six. Once the peer judgment task was completed, participants in the two public text-evaluation conditions proceeded to a post-study questionnaire designed to probe for suspicion.

[Table 3 about here]

The study was a three-condition, between-subject design. Participants were randomly assigned to the “private evaluation” (baseline), “public evaluation/private sanction,” or the “public evaluation/public sanction” condition.¹⁵ We had two main predictions. First, we predicted that participants would rate the text significantly higher in public than in private, indicating the effect of social pressure to conform to the normative evaluation despite private disbelief. Second, our primary prediction was that the more participants conformed by rating the text positively, the more they would enforce the normative evaluation by favoring the four conformist confederates over the lone deviant

¹⁴ As a manipulation check, we asked participants in a post-experimental questionnaire which participant in their group expressed the most divergent opinions from the rest of the group. All but three participants correctly identified participant 6 as the most deviant group member; two of these three gave even more divergent ratings than the lone deviant and correctly identified themselves as most deviant. Thus, participants in our study were almost unanimously aware of which participant was most deviant from the group norm.

¹⁵ Because we were most interested in sanctioning behavior in the study, sixteen participants were assigned to the private evaluation condition and sixty to the other two conditions.

when peer judgments were given in public. However, we do not predict false enforcement among any participants when peer judgments are made privately.

RESULTS

As predicted, participants who evaluated the text publicly tended to rate it higher ($M = 2.73$, about a “B-” average) than those who evaluated it privately ($M = 1.56$, about a “C-/D+” average) ($t(74) = 6.57, p < .001$). This indicates that, in the absence of social pressure, participants recognized that the text was indeed unintelligible. Given that social pressure induced participants to conform to a more favorable normative evaluation than they would otherwise have given, we next test whether social pressure could also induce participants to enforce the norm by favoring conformists over the lone deviant in their publicly submitted peer judgments.

To test for false enforcement, we followed the same approach as in Study 1, analyzing the peer judgments in a condition (public or private) by target (deviant or conformist) by degree of participant’s conformity (text evaluation) repeated-measures ANOVA. Because the text evaluation task did not involve comparison of multiple stimuli (as in Study 1 with the three wine samples), we used the raw, “overall” text evaluation score as the measure of conformity. We treated each participant’s rating of each target (either the lone deviant or one of the four conformists) as an observation. Since each participant in the public text-evaluation conditions rated five targets, the repeated measures dataset consists of 300 observations.

Results are given in Table 4. The overall pattern of effects is similar to the results of Study 1, with the exception that participant’s conformity was also significant,

indicating that participants who gave higher text ratings tended also to generally give higher ratings to all five targets. As in Study 1, we again found a two-way interaction of participant's conformity and target, indicating that the more strongly participants conformed, the higher they rated the conformist targets relative to the deviant target ($F(3, 234) = 17.15, p < .001$). Most important, as in Study 1, there was a significant three-way interaction of deviant target, participant's conformity, and public-private condition. This represents the additional effect of conformists' peer judgments being submitted in public instead of private ($F(1, 234) = 6.72, p = .01$). This 3-way interaction effect indicates that the more participants conformed, the more they favored conformist targets over the deviant, and this effect of conformity was stronger when enforcement was public.

[Table 4 about here]

Figure 2 reports enforcement behavior in Study 2, using a dichotomous instead of continuous measure of conformity for illustration. Those with an "A" or "B" overall rating of the text (similar to that of the four confederates) are labeled "conformists" and those with a "C" or "D" overall rating (similar to that of the deviant) are labeled "non-conformists." We expected those who did not conform to the group consensus would also not feel the need to falsely enforce by sanctioning the lone deviant. This is confirmed in Figure 2. Non-conformists favored the deviant in both conditions, especially in public.

[Figure 2 about here]

Unlike non-conformists, those who conformed were expected to falsely enforce by sanctioning the lone deviant, but only when enforcement was public. Figure 2 shows that when enforcement was private, conformist participants gave about the same

judgments of the conformist as the deviant targets in private, despite having themselves conformed to the normative text evaluation. The critical pattern in Figure 2 is in the public sanctioning condition. While non-conformist participants became more extreme in favoring the lone deviant in public, conformists moved in the opposite direction, rating fellow conformists higher – and the deviant lower – when peer judgments were public. These results provide clear support for our prediction that conformists would give unfavorable peer judgments to deviants, compared to conformist targets, only when evaluations are made in public.

DISCUSSION OF STUDY 2

Study 2 was designed to replicate the results of Study 1 in a different setting in order to demonstrate the robustness of the false enforcement effect. The setting used in Study 2 differed from that of Study 1 in several ways, including an absolute rather than relative evaluation, an erroneously positive instead of erroneously negative consensus, and an analytic instead of aesthetic domain. Study 2 also included an additional “private evaluation” condition for assessing participants’ private beliefs about the text. In the absence of social pressure, participants reported that the text was unintelligible.

In the second part of Study 2, participants again exhibited the predicted pattern of false enforcement found in Study 1. The more participants conformed to the group norm in their text ratings, the more negatively they judged the deviant target relative to the conformist targets – but only when peer judgments were given publicly. This indicates that social pressure not only induced participants to agree with the group consensus, it also induced them to enforce it.

The results of these studies are consistent with our theoretical reasoning that group members, having already conformed to a norm, will also enforce it to signal to other group members the sincerity of their conformity. However, it remains an open question whether this strategy is effective. Does the enforcement of group norms signal the sincerity of conformity more than mere conformity would on its own?

STUDY 3: FROM FALSE ENFORCEMENT TO FALSE SINCERITY

*“The secret of success is sincerity.
Once you can fake that you've got it made.”
-- Jean Giraudoux*

The results of Studies 1 and 2 are consistent with the claim that false enforcement is motivated by the need to publicly affirm the sincerity of false conformity. However, we do not know from these results whether participants in these studies thought that enforcement makes one more likely to be perceived as a “true believer.” Presumably, those who falsely enforce know that their own enforcement behavior is a posture, just like their compliance. Hence, they might assume that everyone else is doing the same. If enforcement is widely regarded as a telltale sign of personal insecurity, social anxiety, and conformity, then it is unlikely that people will enforce to prove to others that they are true believers. Additionally, if false enforcement is viewed as posturing, it may not contribute to overestimation of popular support for the norm, undermining the self-reinforcing dynamic of unpopular norms proposed by Centola et al. (2005).

Study 3 was designed to investigate the credibility of norm enforcement as a signal of sincere conformity. Participants were asked to consider the responses of a protagonist

who had participated in a past study of “text evaluations in groups.” Participants were told that the protagonist expressed a view that an unintelligible academic text by a high status academic was both clear and of high quality, a publicly stated opinion that was very likely a response to group pressure. Participants were told that the protagonist either did, or did not, negatively evaluate another group member who disagreed, depending on the condition of the study. We then asked participants to estimate the protagonist’s privately reported opinion of the text that had been previously measured through a confidential survey. Our theory predicts that participants will assume that a protagonist who had publicly criticized someone for disagreeing with the group’s normative rating of the text is more sincere in their positive opinion of the text, as compared with an otherwise identical protagonist who said nothing critical about others.

METHOD

Forty-five undergraduate students (30 women, 15 men) at a large, private university participated in the study in return for \$8 or extra credit in a sociology class.¹⁶ The study featured a 2 condition (target did or did not enforce) between-subjects design.

Participants were recruited either by fliers advertising payment for participation in a sociology study or via an announced extra credit opportunity in their sociology class.

Participants were directed to a Web page where they signed up to participate in a

¹⁶ Gender of participants did not significantly interact with any of the results presented and is not discussed further.

“Miscellaneous Surveys Study.” Upon arriving in the lab, participants completed a series of unrelated surveys, including a “social perception task.”

The social perception task was presented as a test of the participant’s ability to accurately infer the responses of people involved in an earlier study. In fact the earlier study was entirely fictitious. We designed the study to test whether enforcement behavior would be interpreted as an accurate signal of private beliefs. Participants were given the following introductions:

We recently conducted a laboratory experiment about the evaluation of academic discourse. We had a group of participants read a short essay which you are also going to read in just a few moments. After reading the essay, the participants filled out a questionnaire about their impressions of the text and then put their anonymous questionnaire in a cardboard drop-box. The participants were told that their responses on the questionnaire were completely confidential. None of the other participants ever saw anyone else’s responses. After all the questionnaires were placed in the drop box, we held a group discussion in which the participants publicly discussed their thoughts about the essay with one another. Finally, following the discussion, the participants were told to rate one another’s competence at evaluating the texts. Participants were told that their ratings of the competence of other participants were to be written down on a score card and the scores would then be read to the group by the experimenter.

We are going to show you what one of the participants had to say in public about the essay and about the competence of the other participants. We will then ask you to guess what this participant said earlier about the essay in the anonymous questionnaire that was submitted prior to the group discussion.

After reading this cover story, participants were asked to read the short essay the previous participants had evaluated. We used the same unintelligible text used in Study 2 (Sokal 1996), again identifying the author as “Dr. Robert Nelson, Ph. D., Albert W. Newcombe Professor of Philosophy, MacArthur ‘Genius Grant’ Recipient, 2001, Harvard University.” However, in Study 3, we were not interested in whether these participants

could be pressured to enforce positive evaluations of this text. Instead, we wanted to know if they would interpret public criticism of someone who negatively evaluated the text as an indication that the person making the criticism sincerely believed that the text warranted a positive evaluation.

After reading the essay, participants in Study 3 were asked to review public evaluations of the text submitted by three of the participants in the earlier (fictitious) study, identified as “Participant A, B, and C.” Participants were then shown A’s rating of the competence of “Participant B” and “Participant C.” Participants were told that A and B had both provided “consistently positive comments” in the group discussion, while C had provided “sharply critical comments.” Participants were then told that after submitting their ratings of the text, all group members submitted their ratings of one another on scales ranging from 0 (very low competence) to 10 (very high competence), and that these ratings were then announced to the group. Participants were only given A’s ratings of other group members.

Half of the participants (chosen randomly) were told that A had given B and C ratings of 8 and 3, respectively. The other half were told that A had given both B and C ratings of 5 on the ten-point competence scale. Participants were then asked to guess how A had previously rated the text confidentially, before having heard the views of Participants B and C. (Participants were reminded that A had submitted a private evaluation of the text before participating in the group discussion, using an anonymous questionnaire that was placed in a drop-box). Participants were asked to guess A’s anonymous responses to four items on the anonymous questionnaire that asked A to rate

the text's intelligibility, overall quality, quality of arguments, and understandability on 10-point scales. Each participant's four guesses about A's private text ratings were later summed as a composite measure of the participant's underlying estimate of A's private belief. A reliability analysis of the four items justified construction of the composite measure (Cronbach's alpha = .89).

After this, participants were asked to give their own evaluations of the text using the same scales and items. Finally, after completing the "social perception task," participants were administered a post-study questionnaire. The survey instructions asked participants to think back to the answers they had given earlier and indicate the relative importance of several factors in formulating their guesses about A's true opinion of the text. On 7-point scales,¹⁷ participants indicated the importance of each of the following factors: "His/her ratings of the others' rating abilities," "Suspicion that he/she might be pretending to like the text," and "Your own opinion of the text."

DEBRIEFING

After completing the questionnaire, participants were individually debriefed about the actual goals of the research. A funnel debriefing procedure was used to assess levels of suspicion of deception in the study. No participants reported suspicion. Participants were

¹⁷ Scale formats and the formatting used on each material in the Miscellaneous Surveys study were varied to prevent suspicion that the previous study had been invented solely for the purposes of the Miscellaneous Surveys Study.

assured that their responses were anonymous and that they could request a copy of the research findings.

As a manipulation check, we tested to make sure that participants assigned to each condition did not significantly differ in their evaluations of the text, either because of a failure of random assignment or an unanticipated by-product of the manipulation. No such differences across the two conditions were found, ruling out the possibility that the hypothesized differences between the conditions might be confounded with differences in participants' own ratings of the text.

RESULTS

The main hypothesis in Study 3 was that participants in the “target enforced” condition would expect A’s private opinion of the text (submitted anonymously) to be more positive than would participants in the “no enforcement” condition. Results supported this prediction. Participants in the “target enforced” condition ($M = 7.73$) assumed that A privately liked the text more than did participants in the “no enforcement” condition ($M = 6.31$; $t(45) = 3.24$, $p = .002$). This finding suggests that enforcing a view is effective at increasing the apparent sincerity of that view, even if the enforcement occurs in a setting where suspicion of conformity and false enforcement would be very plausible.

It is interesting to note that participants’ own evaluations of the text had no effect on their estimates of A’s private evaluation ($r = .09$, *ns*), and this did not vary by condition. Thus, even if participants had negatively evaluated the unintelligible text, they did not suspect the target of false enforcement.

Analysis of post-study questionnaire items offered further support for the hypothesis. Participants who were told that A enforced a positive evaluation of the text reported less suspicion that A might have pretended to like the text ($M = 3.65$) than did participants who were told that A did not enforce a positive view ($M = 5.36$) ($t = 3.20, p = .003$). In addition, a t-test showed that the mean reported suspicion in the no-enforcement condition ($M = 5.36$) was significantly higher than the midpoint of the scale ($t = 3.69, p = .001$). Thus, in the experimental setting, not enforcing a positive evaluation of the text led to significant suspicion that the target might have been pretending to like the text.

DISCUSSION OF STUDY 3

An important implication of Studies 1 and 2 is that we should not necessarily assume that enforcers believe the views they pressure others to adopt. If people enforce norms for the same reason they comply, then the enforcement of norms could, ironically, be a sign of posturing. Over time, we might expect people to learn to see through the bluster of self-appointed thought police and to heavily discount the zealotry of those seeking to prove their commitment.

However, Study 3 suggests that this lesson may not be so easy to learn. For every unpopular norm, there are presumably many more norms that are enforced by those who truly believe in them. This allows unpopular norms to “piggyback” on popular norms which reinforce the credibility of enforcement as a signal of sincerity. Study 3 supports such a conclusion. Participants were more likely to assume that a protagonist truly liked the text when they were told that the protagonist gave more favorable scores to those who

also liked the text than to those who didn't. The results suggest that enforcement of a norm is an effective way to affirm the sincerity of one's conformity as a true believer and to avoid suspicion that one has conformed in order to gain approval.

A possible limitation of Study 3 is that participants did not themselves engage in false conformity and therefore did not experience first-hand the motivation to use enforcement to signal the sincerity of compliance. Had they done so, they might have recognized the possibility that enforcement need not reflect private belief. However, research on herd behavior and pluralistic ignorance provides very convincing evidence that people do not project their own motivations onto others (e.g. Miller and McFarland 1991; Prentice and Miller 1993), and research on the "correspondence bias" indicates that people tend to attribute private beliefs to others that are consistent with their observed public behavior (Jones and Harris 1967; Gilbert and Malone 1995). People follow others because they assume others know better or represent the consensus, without considering the possibility that the others they are following are doing the same thing. Nevertheless, future research is needed to determine if this cognitive blind spot applies not only to herd behavior and pluralistic ignorance but also to false enforcement.

GENERAL DISCUSSION AND CONCLUSION

"If there is anything the nonconformist hates worse than a conformist, it's another nonconformist who doesn't conform to the prevailing standard of nonconformity."

-- Bill Vaughan

Prevailing sociological theory assumes that norms emerge because they are useful, whether to society at large, dominant classes, or to individuals with a regulatory interest

in controlling others' behavior. Theories of herd behavior and pluralistic ignorance challenge the traditional view. Both theories show how unpopular norms can be an equilibrium based on an over-estimation of support for a norm. However, the predicted equilibrium is highly unstable because there is motivation only to conform but not to enforce. Thus, people leave the herd as soon as they discover that the beliefs are false, a result anticipated by Andersen in his fable of the naked emperor. The equilibrium collapses when a single child violates the norm with impunity.

However, there are many examples of unpopular norms where the pressure to conform is painfully real. The theory of false enforcement goes beyond herd behavior and pluralistic ignorance to explain why. False beliefs become much more stable when backed up by expectations of enforcement that are confirmed when someone deviates.

A previous study used computational experiments to show how unpopular norms might become self-enforcing. Following Kuran (1995), Centola et al. (2005) used a simple agent-based model to show how a norm with almost no initial support could snowball into a powerful apparent majority when those who conform pressure others to conform as well. The present research tested key behavioral assumptions of the computational model. Study 1 used a variant of the Asch experiment in which participants made judgments about wines instead of lines. However, the main purpose was not to test whether people would conform to a view they did not endorse, but to see if those who conform would also falsely enforce. Had this been the Andersen fable, the lone deviant would have triggered the collapse of the norm. Study participants would have praised the lone deviant, and criticized the four confederates. But that is not what

happened. Studies 1 and 2 showed that participants were not only willing to conform to a normative evaluation they knew to be wrong, but would also publicly criticize a deviant who was praised in the private condition.

This result poses another puzzle: Why would people enforce a norm that they do not want others to obey? We proposed a simple mechanism that can account for false enforcement. The “illusion of sincerity” hypothesis asserts that people enforce norms for the same reason that they conform – as a way to affirm the sincerity of opportunistic compliance. According to this reasoning, false enforcement should only be observed among conformists and only when enforcement is public. Because of the difficulties involved in unambiguously detecting false conformity and enforcement, we tested the hypothesis using controlled laboratory methods. The results of Studies 1 and 2 were consistent with the differences in public and private enforcement we hypothesized based on the “illusion of sincerity” mechanism, though future research should pursue even more direct tests of this mechanism.

Study 3 took an important step in that direction by investigating the effects of false enforcement on perceptions of the sincerity of someone who claimed to appreciate the brilliance of an unintelligible text (a naked emperor familiar to many academics). Those who also enforced agreement with the normative view were regarded as more sincere than were those who did not enforce the norm. This result illuminates the effectiveness of false enforcement in creating an illusion of sincere conformity to a norm. This is important for two reasons. If those who enforce are regarded as “true believers,” 1) this may explain why conformists choose to falsely enforce, and 2) widespread enforcement

may reinforce overestimation of the distribution of popular support for the norm, thereby increasing the pressure to conform. Taken together, the three studies provide support for the theory that unpopular norms are driven by a self-reinforcing dynamic of social pressure-induced enforcement and enforcement-induced social pressure.

In both Studies 1 and 2 we found that the view publicly enforced by participants who had earlier conformed appeared to be inconsistent with their private beliefs. We showed this in two steps. First, we assessed participants' perceptions in the absence of social pressure. In Study 1, a separate test of the stimuli in which participants gave anonymous wine evaluations confirmed that Wines A and B were perceived to be about equal in quality and much better than Wine C which had been tainted with vinegar. In Study 2 we ran an additional experimental condition in which participants gave their evaluations of the academic text anonymously. Again, the results confirmed that participants' perceptions were inconsistent with the group norm, as they gave low ratings to the unintelligible academic text.

Second, we compared public and private peer judgments for further evidence of participants' private beliefs. In both Studies 1 and 2, the three-way interaction showed that conformity affected peer judgments in public but not in private. In private, it appears that participants expressed their genuine regard for the lone deviant, which suggests private agreement. It could also indicate admiration for and tolerance of dissent (Moscovici and Nemeth 1974), but this is qualitatively different from the conviction that the norm is true – that this is what people should believe. The litmus test of this conviction is private enforcement, because this indicates sufficient certainty about one's

beliefs that one disapproves of those who disagree. Those who fail to privately enforce – because of admiration for and/or tolerance of dissent – lack the conviction that others should believe as they do. Should they nevertheless enforce the norm in public, that enforcement is false.

False enforcement appears to require some level of uncertainty. A pilot version of Study 1 in which confederates established a norm favoring the tainted wine showed that, when participants were certain the normative view was wrong, the group consensus had insufficient credibility to induce conformity. It is therefore important to recognize that false enforcement is unlikely to occur in all situations, especially those where uncertainty regarding the norm is quite low. For example, in Jonestown it is very possible that conformity to, and enforcement of, the norm to drink the Flavor Aid might not have occurred if the repeated “white nights” had not made the fatal consequences of the ritual unclear to Jones’ followers.

In addition to uncertainty regarding the norm, other factors may also be necessary for the false enforcement results we obtained. In the present research there were no explicit costs to enforcement. Had sanctioning been costly, the rates of false enforcement would likely have diminished. However, costly enforcement might also be a more effective signal of sincerity than simply submitting a negative peer judgment, which might be regarded by skeptics as “cheap talk” (Frank 1988). Further research is needed to see how false enforcement is affected by the cost of sanctioning.

It is possible that underlying beliefs of participants in our studies changed in the direction of the norm, either as a result of “informational influence” (Deutsch and Gerard

1955) or as a consequence of cognitive dissonance reduction following conformity (Festinger 1954). Recent neurobiological research indicates that perception can be shaped by social influence (Berns et al. 2005), which suggests the possibility that participants may have come to believe that Wine B really was inferior and the text was intelligible. However, it is important to keep in mind that those who defer to the majority because the consensus has undermined their confidence that the norm is false are very different from people who agree with the consensus because of their conviction that the norm is true. The private wine and text ratings, and the absence of communication in our experimental design (other than submission of one's ratings), make it unlikely that participants shifted that far in their beliefs. Although informational influence and cognitive dissonance reduction cannot account for the differences we observed between public and private peer judgments, we cannot rule out the possibility that conformity and false enforcement could lead to underlying belief change in natural settings. Indeed, Centola et al. (2005) included this possibility in their computational experiments and found, surprisingly, that this led to the collapse of the norm, rather than to reinforcement.

For Studies 1 and 2, we used a wine tasting and an academic text evaluation. Consensus judgments about the quality of wines or academic texts may not be what we usually think of as norms. We usually think of norms as implying "should" statements regarding appropriate attitudes and behavior, but pressure to agree with and enforce an unwarranted social judgment corresponds with a class of norms that are all too familiar. These include judgments intended to express cultural or intellectual sophistication, group superiority, adolescent independence, or political patriotism. Further, substantial theory and research suggests that violations of accepted convention in culturally and

intellectually sophisticated environments can lead to shame and ridicule of those exposed as lacking sufficient cultural credentials (Bourdieu [1979]1984; Elias [1939]1978; Goffman 1963). It is not surprising that people go along with social judgments they privately do not endorse, given the social pressure. What is harder to explain is how this social pressure becomes self-reinforcing. That was the question these studies were intended to help us answer, though under less heavily charged and more controlled conditions.

It is important to keep in mind that the participants in Studies 1 and 2 were unrelated individuals and that the study was conducted under the minimal conditions of spontaneously formed groups of strangers interacting via computer terminals with limited expectation of future interaction with one another. The false enforcement observed in these studies might have been even more pronounced if participants were members of a group on which they depended for social approval (Hechter 1987). Religious, military, and professional groups that are able to demand high levels of sacrifice from their members may be especially prone to unpopular norms because of greater anxiety about being regarded as insufficiently sincere – a “poser” who conforms simply to gain approval. Future research needs to be directed to these groups with “strong solidarity” (Lindenberg 1988), to norms with political, religious, or ideological content, and to “anachronistic norms” that persist because of social control despite having become dysfunctional (Ogburn 1922; Triandis 1994; Vandello and Cohen 2004).

Another avenue for future research is to examine the motivations for false enforcement. Study 3 shows that public enforcement is an effective way to appear

sincere, but more research is needed to see if false enforcers were consciously trying to create the “illusion of sincerity.” Ethnographic research could explore the dynamics of honor systems based on explicit metanorms that assign equal guilt to those who fail to report violations of the honor code. As discussed above, another possible motivation for false enforcement is that acting like a “true believer” reduces cognitive dissonance among those who falsely conform. Alternatively, one could interpret false enforcement as a manifestation of “reaction formation” (Freud [1898]1962) by those who feel shame for their private deviance. However, the differences we observed between public and private enforcement are difficult to reconcile with alternative explanations based on dissonance reduction and reaction formation, whereas our account based on the illusion of sincerity fits well with this pattern.

Finally, in both Studies 1 and 2 the lone deviant was lower status than the conformist targets to bolster the credibility of the normative wine and text ratings. In addition, participants in Study 2 were asked to evaluate a text authored by a prestigious academic. Because these features were controlled across the public and private conditions they do not present alternative explanations for the differences in public and private peer judgments observed in both studies. Nevertheless, the high status of “true believers” may be essential for false enforcement for two reasons: 1) to induce public conformity with a norm that is privately questioned, and 2) to increase the need to prove sincerity in order to secure approval from those whose judgments carry great weight. More research is needed to investigate the interactions between status standing and norm enforcement. While we encourage pursuit of a wide range of alternative theories and expect that there are likely to be several ways that false enforcement can be generated, we believe explicit

metanorms and the need to demonstrate sincerity are the most promising explanations so far.

One type of status difference that might be especially interesting to examine is that of the new recruit who seeks acceptance and approval from the established members of the group. In extreme cases, new recruits to a group might become the most fervent enforcers of norms they privately question, sanctioning others out of a need to prove their own commitment. Ironically, the “true believers” who conform out of genuine conviction may find that they can leave enforcement to these zealots, in which case enforcement will be a reliable indicator of private doubt, not true belief. Although participants in Study 2 did not embrace this possibility, if the “illusion of sincerity” hypothesis is correct, then an effective counter-strategy might be to persuade group members that over-zealous norm enforcement is a sign of insecurity, not conviction. As Stanley Milgram (1974:568) noted, “It may be that we are puppets – puppets controlled by the strings of society. But at least we are puppets with perception, with awareness. And perhaps our awareness is the first step to our liberation.”

Table 1: “Overall” ratings of wines submitted by simulated confederates in Study 1

	Participant 1	Participant 2	Participant 3	Participant 4	Participant 6
Wine A	A	A	B	A	A
Wine B	D	C	D	D	A
Wine C	C	C	C	C	C

Table 2: Results of repeated-measures analysis of variance, with condition (public or private), participants' conformity (participant's Wine A rating minus Wine B rating), and target (deviant or conformist) predicting participants' peer judgments of the simulated confederates in Study 1.

Variable	<i>F</i>	d.f.
<i>Between-subjects</i>		
Public Enforcement	7.44**	1
Participant's Conformity	1.54	3
Public x Participant's Conformity	1.58	3
<i>Within-subjects</i>		
Deviant Target	1.61	1
Deviant Target x Public	1.41	1
Deviant Target x Participant's Conformity	2.85*	3
Deviant Target x Participant's Conformity x Public	3.16*	3
Residual		196
* $p < .05$, ** $p < .01$. All tests are two-tailed.		

Table 3: “Overall” ratings of text submitted by simulated confederates in Study 2

Participant 1	Participant 2	Participant 3	Participant 4	Participant 6
A	A	A	A	D

Table 4: Results of repeated-measures analysis of variance, with condition (public or private), participants' conformity (participant's text rating), and target (deviant or conformist) predicting participant's peer judgments of the simulated confederates in Study 2.

Variable	<i>F</i>	d.f.
<i>Between-subjects</i>		
Public Enforcement	0	1
Participant's Conformity	4.94**	3
Public x Participant's Conformity	2.65	1
<i>Within-subjects</i>		
Deviant Target	3.8	1
Deviant Target x Public	0.03	1
Deviant Target x Participant's Conformity	17.15**	3
Deviant Target x Participant's Conformity x Public	6.72*	1
Residual		234
* $p < .05$, ** $p < .01$. All tests are two-tailed.		

Figure 1: Peer judgments in Study 1 by condition (public or private enforcement), target (conformist or deviant) and participant's conformity. Participant "conformed" denotes an evaluation of Wine A > Wine B; "did not conform" denotes Wine A ≤ Wine B.

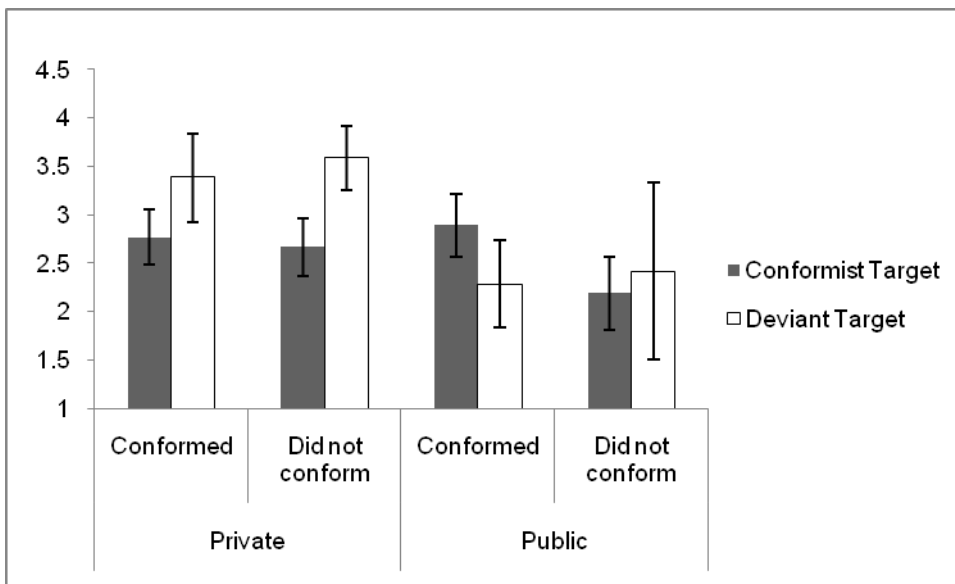
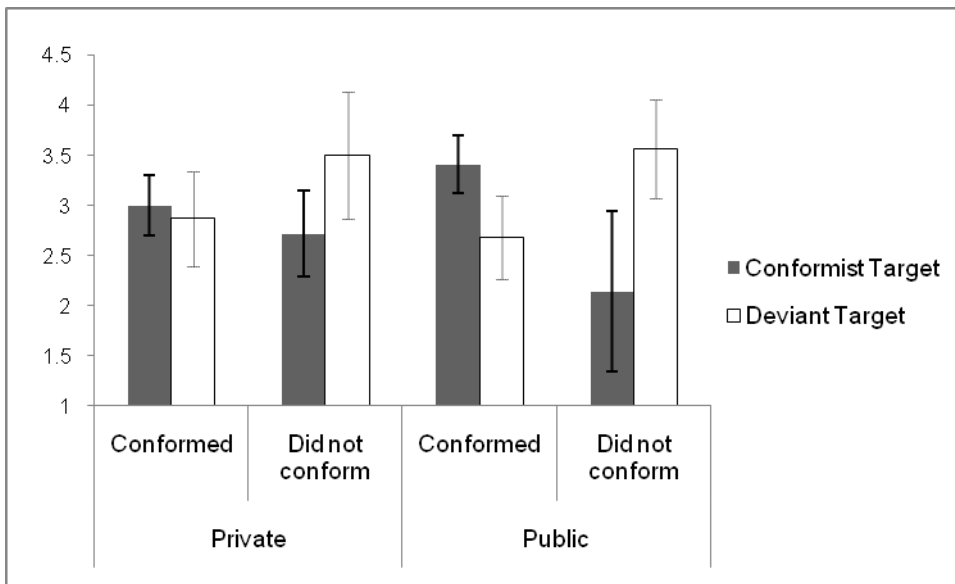


Figure 2: Peer judgments in Study 2 by condition (public or private enforcement), target (conformist or deviant) and participant's conformity. Participant "conformed" denotes an evaluation score of A or B; "did not conform" denotes C or D.



REFERENCES

- Aguinis, Herman and Mitchell M. Handelsman. 2006. "The Unique Ethical Challenges of the Bogus Pipeline Methodology: Let the Data Speak." *Journal of Applied Social Psychology*. 27: 582-587.
- Andersen, Hans Christian. [1837](1998). *Tales*. Vol. XVII, Part 3. *The Harvard Classics*. New York: P.F. Collier & Son.
- Arrow, Kenneth J. 1971. "Political and Economic Evaluation of Social Effects and Externalities," edited by M. D. Intriligator. *Frontiers of Quantitative Economics*. Amsterdam: Elsevier Science.
- Arthur, Brian. 1990. "Positive Feedbacks in the Economy." *Scientific American*. February: 92-99.
- Asch, Solomon E. 1951. "Effects of Group Pressure Upon the Modification and Distortion of Judgments." Pp. 177-90 in *Groups, Leadership, and Men*, edited by H. Guetzkow. Pittsburgh: Carnegie Press.
- Asch, Solomon E. 1955. "Opinions and Social Pressure." *Scientific American*. 193:33-35.
- Axelrod, Robert. 1986. "An Evolutionary Approach to Norms." *American Political Science Review*. 80(4), 1095-1111.

- Baer, J. S. 1994. "Effects of College Residence on Perceived Norms for Alcohol Consumption: An Examination of the First Year in College." *Psychology of Addictive Behaviors*. 8:43–50.
- Banerjee, Abhijit. 1992. "A Simple Model of Herd Behavior." *The Quarterly Journal of Economics*. 107: 797-817.
- Berns, Gregory S., Jonathan Chappelow, Caroline F. Zink, Giuseppe Pagnoni, Megan E. Martin-Skurski, and Jim Richards. 2005. "Neurobiological Correlates of Social Conformity and Independence During Mental Rotation." *Biological Psychiatry*. 58: 245-253.
- Bicchieri, Cristina and Y. Fukui. 1999. "The Great Illusion: Ignorance, Informational Cascades and the Persistence of Unpopular Norms." *Business Ethics Quarterly* 9: 127-155.
- Bikhchandani, Sushil, David Hirshleifer, and Ivo Welch. 1992. "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades." *Journal of Political Economy*. 100: 992-1026.
- Bikhchandani, Sushil, David Hirshleifer, and Ivo Welch. 1998. "Learning from the Behavior of Others: Conformity, Fads, and Informational Cascades." *Journal of Economic Perspectives*. 12: 151-170.
- Bond, R. and P. B. Smith. 1996. "Culture and Conformity: A Meta-Analysis of Studies Using Asch's (1951, 1956) Line Judgment Task." *Psychological Bulletin*. 119:111-137.

- Borsari, Brian and Kate B. Carey. 2001. "Peer Influences on College Drinking: A Review of the Research." *Journal of Substance Abuse* 13: 391–424.
- Bourdieu, Pierre. [1979]1984. *Distinction: A Social Critique of the Judgment of Taste*. London: Routledge.
- Bowles, Samuel and Herbert Gintis. 2001. "Social Capital and Community Governance." Working Paper. Santa Fe Institute.
- Centola, Damon, Robb Willer, and Michael W. Macy. 2005. "The Emperor's Dilemma: A Computational Model of Self-Enforcing Norms." *American Journal of Sociology*. 110(4):1009-40.
- Chen, Yehning. 1999. "Banking Panics: The Role of the First-Come, First-Served Rule and Information Externalities." *Journal of Political Economy*. 107(5):946-968.
- Cialdini, Robert B. 1988. *Influence: Science and Practice*, 2nd edition. Boston: Scott, Foresman and Co.
- Clance, P. R. 1985. *The Impostor Phenomenon: Overcoming the Fear that Haunts Your Success*. Atlanta: Peachtree Publishers.
- Deutsch, Morton and Harold B. Gerard. 1955. "A Study of Normative and Informational Social Influences Upon Individual Judgment." *Journal of Abnormal and Social Psychology*. 51:629-36.
- Elias, Norbert. [1939]1978. *The Civilizing Process, Vol.I. The History of Manners*. New York: Pantheon.

- Fehr, Ernst and Rus Fischbacher. 2004. "Third-Party Punishment and Social Norms." *Evolution and Human Behavior*. 25: 63-87.
- Fehr, Ernst and Simon Gächter. 2000. "Cooperation and Punishment in Public Goods Experiments." *American Economic Review*. 90: 980-994.
- Fehr, Ernst and Simon Gächter. 2002. "Altruistic Punishment in Humans." *Nature*. 415: 137-140.
- Festinger, Leon. 1954. *A Theory of Cognitive Dissonance*. Stanford, CA: Stanford University Press.
- Frank, Robert H. 1988. *Passions Within Reason: The Strategic Role of the Emotions*. New York: W.W. Norton and Co.
- Freud, Sigmund. [1898] 1962. "The Neuro-Psychoses of Defense." *The Standard edition of the Complete Psychological Works of Sigmund Freud, Vol. 3*. London: Hogarth Press.
- Gaes, Gerald G., Robert J. Kalle, and James T. Tedeschi. 1978. "Impression Management in the Forced Compliance Situation." *Journal of Experimental Social Psychology*. 14:579-587.
- Gilbert, Daniel T. and Patrick S. Malone. 1995. "The Correspondence Bias." *Psychological Bulletin*. 117: 21-38.

- Gilovich, Thomas, Kenneth Savitsky, and Victoria Husted Medvec. 1998. "The Illusion of Transparency: Biased Assessments of Others' Ability to Read One's Emotional States." *Journal of Personality and Social Psychology*. 75: 332-346.
- Gilovich, Thomas, Victoria Husted Medvec, and Kenneth Savitsky. 2000. "The Spotlight Effect in Social Judgment: An Egocentric Bias in Estimates of the Salience of One's Own Actions and Appearance." *Journal of Personality and Social Psychology*. 78:211-222.
- Goffman, Erving. 1963. *Stigma: Notes on the Management of Spoiled Identity*. Englewood Cliffs, NJ: Prentice-Hall.
- Graham, John R. 1999. "Herding Among Investment Newsletters: Theory and Evidence." *Journal of Finance*. 54:237-268.
- Gross, Paul R. and Norman Levitt. 1994. *The Academic Left and its Quarrels with Science*. Baltimore: Johns Hopkins.
- Hanson, Ward A. and Daniel S. Putler. 1996. "Hits and Misses: Herd Behavior and Online Product Popularity." *Marketing Letters*. 7:297-306.
- Harray, Keith. 1992. "The Truth about Jonestown: 13 years later - why we should still be afraid." *Psychology Today*. March-April.
- Hechter, Michael and Karl-Dieter Opp. 2003. "Introduction." Pp. xi-xx in *Social Norms*, edited by Michael Hechter and Karl Dieter-Opp. New York: Russell Sage.

- Hechter, Michael. 1987. *Principles of Group Solidarity*. University of California Press: Berkeley.
- Heckathorn, Douglas D. 1988. "Collective Sanctions and the Creation of Prisoner's Dilemma Norms." *American Journal of Sociology*. 94(3): 535-562.
- Hirshleifer, David. 1995. "The Blind Leading the Blind: Social Influence, Fads, and Informational Cascades." In *The New Economics of Human Behavior*, (eds. Mariano Tommasi and Kathryn Ierulli). Cambridge, UK: Cambridge University Press.
- Horne, Christine. 2001a. "Sociological Explanations of the Emergence of Norms." Pp. 3-34 in *Social Norms*, edited by Michael Hechter and Karl Dieter-Opp. New York: Russell Sage.
- Horne, Christine. 2001b. "The Enforcement of Norms: Group Cohesion and Meta-Norms." *Social Psychology Quarterly*. 64(3): 253-266.
- Horne, Christine. 2004. "Collective Benefits, Exchange Interests, and Norm Enforcement." *Social Forces*. 82(3): 1037-1062.
- Horne, Christine. 2009. *The Rewards of Punishment*. Stanford University Press. Palo Alto, CA.
- Jones, Edward E. and Keith Davis. 1967. "The Attribution of Attitudes." *Journal of Experimental Social Psychology*. 3: 1-24.

- Jones, Edward E., and Harold Sigall. 1971. "The Bogus Pipeline: A New Paradigm for Measuring Affect and Attitude." *Psychological Bulletin*. 76: 349-364.
- Krech, David and Richard S. Crutchfield. 1948. *Theories and Problems of Social Psychology*. New York: McGraw-Hill.
- Kuran, Timur. 1995. *Private Truths, Public Lies: the Social Consequences of Preference Falsification*. Harvard University Press: Cambridge, Massachusetts.
- LaTour, Stephen A. and Ajay K. Manrai. 1989. "Interactive Impact of Informational and Normative Influence on Donations." *Journal of Marketing Research*. 26: 327-335.
- Lee, In Ho. 1998. "Market Crashes and Informational Avalanches." *The Review of Economic Studies*. 65(4): 741-759.
- Maaga, Mary M. 1998. *Hearing the Voices of Jonestown*. Syracuse, NY: Syracuse University Press.
- Mackie, Gerry 1996. "Ending Footbinding and Infibulation: A Convention Account." *American Sociological Review*. 61: 999-1017.
- Marx, K. & Engels, F. [1845] 1986. "The German Ideology." Pp. 13-14 in *Politics and Ideology*, edited by J. Donald & S. Hall Milton Keynes: Open University Press.
- Merton, Robert K. 1968. "The Matthew Effect in Science." *Science*. 159: 56.
- Milgram, Stanley. 1974 (October 31). "We Are All Obedient." *The Listener*. pp.567-568.
- Miller, Arthur. 1996. "Why I Wrote *The Crucible*." *New Yorker*, October 21.

- Miller, Dale T., and Cathy McFarland. 1991. "When Social Comparison goes Awry: The Case of Pluralistic Ignorance." P. 287-313 *Social Comparison: Contemporary Theory and Research*, edited by J. Suls & T.A. Wills. Hillsdale, N.J.: Erlbaum.
- Moscovici, Serge, and Charlan Nemeth. 1974. "Social Influence II: Minority Influence." Pp. 217-249 in *Social Psychology: Classic and Contemporary Integrations*, edited by C. Nemeth. Chicago: Rand McNally.
- Nagoshi, G. T., Nakata, T., Sasano, K., & Wood, M. D. 1994. "Alcohol Norms, Expectancies, and Reasons for Drinking and Alcohol Use in a U.S. versus a Japanese College Sample." *Alcoholism: Clinical and Experimental Research*. 18(3): 671-678.
- Noelle-Neumann, Elisabeth. 1984. *The Spiral of Silence: Public Opinion - Our Social Skin*. Chicago: The University of Chicago Press.
- O'Gorman, Hubert. 1986. "The Discovery of Pluralistic Ignorance: An Ironic Lesson." *Journal of the Historic and Behavioral Sciences*. 22:333-347.
- Ogburn, William F. 1922. *Social Change with Respect to Culture and Original Nature*. New York: Huebsch
- Oliver, Pamela E. 1980. "Rewards and Punishments as Selective Incentives for Collective Action: Theoretical Investigations." *American Sociological Review*. 45(6):1356-1375.

- Opp, Karl-Dieter, Christiane Gern and Peter Voss. 1996. *Origins of a Spontaneous Revolution: East Germany*. Ann Arbor: University of Michigan Press.
- Ostrom, Elinor, James Walker, and Roy Gardner. 1992. "Covenants With and Without a Sword: Self-Governance Is Possible." *American Political Science Review*. 86(2): 404-417.
- Parsons. Talcott. 1971. *The System of Modern Societies*. Englewood Cliffs NJ: Prentice-Hall.
- Perkins, H.W. and H. Wechsler. 1996. "Variation in perceived college drinking norms and its impact on alcohol abuse: A nationwide study." *Journal of Drug Issues*. 26: 961-974.
- Prentice, Deborah A. and Dale T. Miller. 1993. "Pluralistic Ignorance and Alcohol Use on Campus: Some Consequences of Misperceiving the Social Norm." *Journal of Personality and Social Psychology*. 64(2):243-56.
- Salganik, Matthew J. Peter S. Dodds, and Duncan J. Watts. 2006. "Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market." *Science*, 311: 854-856.
- Salganik, Matthew J., and Duncan J. Watts. 2008. "Leading the Herd Astray: An Experimental Study of Self-Fulfilling Prophecies in an Artificial Cultural Market." *Social Psychology Quarterly*. 71:338-355.
- Savistky, Kenneth, Nicholas Epley, and Thomas Gilovich. 2001. "Do Others Judge Us As Harshly As We Think? Overestimating the Impact of Our Failures,

Shortcomings, and Mishaps.” *Journal of Personality and Social Psychology*.81:44-56.

Schiller, Robert J. 2000. *Irrational Exuberance*. Princeton, NJ: Princeton University Press.

Schulman, Gary I. 1967. “Asch Conformity Studies: Conformity to the Experimenter and/or to the Group?” *Sociometry*. 30:26-40.

Shinada, Zizuho, Toshio Yamagishi, and Yu Ohmura. 2004. “False Friends Are Worse Than Bitter Enemeis: ‘Altruistic’ Punishment of in-Group Members.” *Evolution and Human Behavior*. 25: 379-393.

Sokal, Alan and Jean Bricmont. 1998. *Fashionable Nonsense: Postmodern Intellectuals’ Abuse of Science*. New York: Picador.

Sokal, Alan. 1996. “Transgressing the boundaries: Towards a Transformative Hermeneutics of Quantum Gravity.” *Social Text* 46/47, 217-252.

Tedeschi, James T., Barry R. Schlenker, and Thomas V. Bonoma. 1971. “Cognitive Dissonance: Private Ratiocination or Public Spectacle?” *American Psychologist*. 26:685-695.

Triandis, Harry C. (1994). *Culture and Social Behavior*. New York: McGraw-Hill.

Vandello, Joseph A. and Dov Cohen. 2003 “Male Honor and Female Fidelity: Implicit Cultural Scripts that Perpetuate Domestic Violence.” *Journal of Personality and Social Psychology*. 84: 997-1010.

- Vandello, Joseph A. and Dov Cohen. 2004. "When Believing is Seeing: Sustaining Norms of Violence in Cultures of Honor." In *Psychological Foundations of Culture*, edited by M. Schaller & C. Crandall. Mahwah, NJ: Erlbaum.
- Willer, Robb. 2004. "The Intelligibility of Unintelligible Texts." Master's Thesis. Cornell University. Ithaca, NY.
- Willer, Robb. 2005. "Overdoing Gender: A Test of the Masculine Overcompensation Thesis." Paper presented at the Annual Meeting of the American Sociological Association. Philadelphia, PA.
- Yamagishi, Toshio. 1986. "The Provision of a Sanctioning System As a Public Good." *Journal of Personality and Social Psychology*. 51: 110-116.
- Yamagishi, Toshio. 1988a. "The Provision of a Sanctioning System in the United States and Japan." *Social Psychology Quarterly*. 51: 265-271.
- Yamagishi, Toshio. 1988b. "Seriousness of Social Dilemmas and the Provision of a Sanction System." *Social Psychology Quarterly*. 51(1): 32-42.