# Essays on Social Norms

By

Minjae Kim

B.A. Political Science
University of Chicago, 2012

S.M. Management Research
Massachusetts Institute of Technology, 2017

SUBMITTED TO THE SLOAN SCHOOL OF MANAGEMENT IN PARTIAL
FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY IN MANAGEMENT

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

JUNE 2018

Signature redacted

Signature of Author:_____
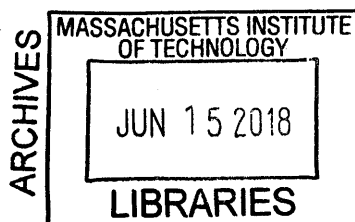Department of Management
April 19, 2018

Signature redacted

Certified by: _____
Ezra W. Zuckerman Sivan
Deputy Dean / Alvin J. Siteman (1948) Professor of Strategy and Entrepreneurship
Thesis Supervisor

Signature redacted

Accepted by: _____
Catherine Tucker
Sloan Distinguished Professor of Management
Professor of Marketing
Chair, MIT Sloan PhD Program

# Essays on Social Norms

by

Minjae Kim

Submitted to the Sloan School of Management on April 19, 2018,
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in Management

ABSTRACT

The first essay addresses why people might conform to norms that they do not endorse. One explanation is *pluralistic ignorance*: when everyone appears to endorse a norm, even non-endorsers will conform so as to feign their commitment to the group's values, thereby exacerbating the misperception. But this explanation is limited because people seem to even conform to norms that are widely known to contradict the group's values ("visibly unpopular norms"), to the point that their conformity appears insincere. I argue that such insincere conformity is an especially potent signal of commitment because it shows that one is willing to sacrifice one's personal preferences on others' behalf. Using both qualitative and experimental methods, I study the visibly unpopular norm prescribing excessive drinking in after-hour business gatherings in South Korea. The analysis indicates that an insincere conformist to the drinking norm is perceived as an especially committed relationship partner. An important implication is that some norms might persist *not despite* the fact they contradict group members' preferences but *because of* this contradiction.

The second essay (coauthored with Ezra W. Zuckerman Sivan) addresses why norms might not persist despite their wide popularity. Recent research suggests that many norms may be upheld by closet deviants who engage in enforcement so as to hide their deviance. But various empirical accounts indicate that audiences are often quite sensitive to this ulterior motive. Our theory and experimental evidence identify when inferences of ulterior motive are drawn and clarify the implications of such inferences. Our main test pivots on two contextual factors: (1) the extent to which individuals might try to strategically feign commitment and (2) the contrast between "mandated" enforcement, where individuals are asked for their opinions of deviance, and "entrepreneurial" enforcement, where enforcement requires initiative to interrupt the flow of social interaction. When the context is one where individuals might have a strategic motive and enforcement requires entrepreneurial initiative, suspicions are aroused because the enforcers could have remained silent and enjoyed plausible deniability that they had witnessed the deviance or recognized its significance. Given that the mandate for enforcement might be rare, a key implication is that norms might frequently be under-enforced.

Thesis Supervisor: Ezra W. Zuckerman Sivan
Title: Deputy Dean / Alvin J. Siteman (1948) Professor of Strategy and Entrepreneurship

# Acknowledgements

As social scientists, we refrain from attributing a driver of some social phenomenon to sheer *luck*. This is because much of our profession's business is by definition generated by efforts to use *systematic* accounts in explaining variance (or lack thereof) in behaviors and other valued outcomes in our social worlds. And as a student of social sciences, I can indeed account for many steps in my life relatively well using various social scientific theories. That is, chances of my being where I am right now, especially with respect to my educational achievement, are relatively high, based on our current understanding of social worlds. I was born in a recently developed country where education is highly prioritized; both of my parents went to college, and there are some PhD's in my extended family; and I was well-fed and supported to participate in a wide range of activities through my primary, secondary, and higher education. While there might be a few things in my life that might be less-likely-than-average to happen, I would not need to rely on a terrible amount of luck to account for steps I have taken from the day I was born up to today.

Thus, from a social scientific point of view, the fact that I have never had to beat the odds to get here means that little of my life outcome is due to luck. By contrast, if we take a more moral point of view, the fact that I have never had to beat the odds means that so much of my life outcome is actually due to luck. That is, if luck is understood as how much I have achieved given how much I *deserve*, I have been incredibly lucky all my life. This realization is especially salient in instances such as those countless walks out of E62, when the usually-bustling Main Street is empty besides a few cabs idling for last passengers and shopkeepers planning their commute home on Saturday nights. That serenity amidst the confused and perplexed thoughts of a novice social scientist. Few people *deserve* that, and I am certainly not one of them.

And none of this incredible luck in my graduate student career can be described without mentioning people who have supported me along the way. One of the guiltiest is certainly my advisor, Ezra W. Zuckerman Sivan. I still remember receiving a phone call from him on April 12th, 2013, in which he said that I was off the waitlist but nevertheless warned me with reasons why I should essentially not come to Sloan as a PhD student. Little did I know back then that that conversation only reflects his care and devotion for students like me for whom he works so hard to support, above and beyond what is typically expected of an advisor. There have since been countless emails exchanged at 3am when I needed his help on finishing a draft and he was relentlessly supporting me but at the same time never lowering standards. It is his stubbornness for high-quality research combined with his genuine care for my intellectual and personal growth that has motivated me to delve deeper into my curiosity throughout my graduate career and will do so for rest of my career. Even as a deputy dean with lots of administrative duties, he never takes more than a few hours to respond to my email (if that), and he always is ready to talk to me about a new crazy idea I have come up with, with seriousness that I do not deserve yet that is always given to me. Even though it would be very difficult to find a more decorated CV than his, people close to him know him for his genuine care and intellectual generosity for others.

Combine that with unending curiosity and youthful energy for insights into our social world, to think of who Ezra is as an academic and an advisor. I can only hope that I can emulate even just half of his stubbornness for quality, generosity for others, and endless curiosity; and I am extremely honored to graduate Sloan holding the title of being his student.

I would never be who I am right now without those walks up and down E62 with Roberto M. Fernandez in countless early mornings of the past five years. From my day one at Sloan, he has been the watchful guardian for my personal as well as intellectual development, and I have never seen an academic as selfless as him (and I doubt that I will see another either). I will never forget the meetings where he would sit down with me to discuss my future as a student and an academic in general: what project will I be working on next and how should I prioritize my time? In discussing these important matters of graduate school career, he has always put my interests ahead of everyone's, including his, paying a great amount of respect to my own opinions and desires. And all this is not to downplay his intellectual might. His (both deep and versatile) insights often clarify points that authors themselves did not know were in their papers, and I have also been a beneficiary of such insights many times. The experience of writing my first academic paper was possible entirely because of him, and both his intellectual guidance and patience with me along the way are reasons why I can think of writing academic papers as both fun and rewarding. When I leave Sloan, I will miss having those walks up and down the building, during which much of my personal as well as intellectual growth took place.

Oliver Hahl and I might have seen each other in person perhaps fewer than 20 times for the past five years, but he has become one of the biggest mentors in my life in general. A half-serious joke is that I saw him give a talk in Chicago before I came to Sloan, and I was so mesmerized by his research that I knew from then that I would want to be his coauthor. And it is amazing how big part of my work as well as personal life he has become. In conversations with him, I can actually feel how much he cares about my success as a person as well as an academic, as he spends so much effort thinking of what is best for me. He always listens and gives incredibly thoughtful advice; and sharpness in his comments is just not to be found anywhere else. Most of all, it has always been tremendously fun to work with him, as he just is a genuinely good person to be around. Although he might not want the title, he will always be a 선생님 for me (note: Oliver speaks Korean), and I can only hope to be able to follow his footsteps as a fellow graduate of Sloan.

I am also lucky to have Kate Kellogg on my committee. Many would be surprised by her being on my committee, given that my research barely speaks to her lines of research. And that is only a testament to her versatility and generosity as a scholar and a teacher. Whenever I was struggling to figure out what it even is that I am talking about, she would help me take a step back and see what is in the idea that I am trying to uncover. I admire how much care she devotes to wellbeing of students, and I have been incredibly lucky to be on the receiving end of it. The Behavioral Policy Sciences community at Sloan is lucky to have her; and I have been a huge beneficiary of that luck for the past five years, which I will miss dearly.

Although Susan Silbey is not part of my official committee, her endless mentorship has been indispensable for my development as a social scientist. Her class on research design basically introduced me to what social science as an intellectual endeavor is, and I only hope to produce research that can live up to the standard I learned in that class and she practices in her teaching and own research. She has also been an incredibly patient teacher, walking me through different methodological, theoretical, and epistemological lines of thought. She does not shy away from learning new things, which I hope to have courage for as much as she does, but which in retrospect is not surprising given her versatility as a scholar. Perhaps most importantly, she has been a steadfast mentor for my personal growth as well, and dinners and chats with her will be among the fondest memories of my graduate school years. Without her beside me for the past five years, I would not be where I am today, and I thank her for her patient care and intellectual generosity.

I owe great thanks to many other faculty as well. As Roberto would say, it takes a village to educate a PhD student, and that has certainly been the case for me. In particular, I thank Ray Reagans, Emilio Castilla, Erin Kelly, JoAnne Yates, Evan Apfelbaum, Olenka Kacperczyk, John Carroll, and Jared Curhan. Their support in class, seminars, and one-on-one meetings, among others, has been essential in each step of my intellectual life at Sloan.

My growth as a social scientist would be incomplete without a wonderful group of fellow students that I have also been lucky to have. They have helped me through thick and thin, everything from helping register for courses to practicing talks together to reading each other's drafts, all of which I could not survive without. Four of those angels are Vanessa Conzon, Rebecca Grunberg, Erik Duhaime, and Brittany Bond. They are the people I can always rely on emotionally as well as intellectually, and my friendship with them alone is worth the five years I have spent here. Vanessa has become someone I can talk about anything, and the countless dinners with her (and her husband Eric Kilpatrick and their amazing daughter, Hazel Conzon) have provided me with comfort of being home whenever I needed it. I thank Rebecca for always challenging things that I take for granted; she makes me realize what assumptions I have, in living everyday life as well as writing papers. Our hiking trips in New Hampshire will be among the fondest memories of my time in graduate school. Erik's youthful energy and constant encouragement always remind me how exciting this intellectual endeavor we are in is. I also thank Brittany for being the reliable companion in our journey together in the program, as I will miss our Friday afternoon chat sessions that range from gossip to deep discussions about the state of our academic field.

I have also been incredibly lucky to have gained four *noona's* (older sisters, in Korean) during my time here. In particular, Jae-Kyung Ha has been the exemplary scholar and person in general that I want to emulate. Without her as a role model in the program, there is no doubt that I would have faltered, and she will forever be my hero for being such a considerate person as well as a sharp researcher. Hyejin Rho and Hye Jun Kim have also been my rocks for the past five years, often feeding me amazing Korean food and teaching me about life in general. Eunhee

Sohn has also been nothing but a caring *noona*, as I always look forward to connecting with her and her husband, Brian Park, whenever possible.

My cohort members have also been nothing but supportive during my journey here. I will never find a better officemate (and a friend) than Sam Zyontz, and neither will I forget those many times when we were the only people on our floor on weekends, pushing each other to keep making progress. Danny Kim has also become my dear friend during our five years together here, not hesitating to give me advice on personal as well as academic stuff when I need it the most. James Riley, Ankur Chavda, and Daniel Rock were also always there for me, giving me reality checks as well as generous emotional support.

My fellow students in the Economic Sociology Program and neighboring programs in the Behavioral Policy Sciences also played a big part in my intellectual and personal development here. In particular, while Michael Wahlen would be too humble to say it himself (even though he probably knows it!), I have always learned more from him than he has from me, even when he asks for help from me. I will forever be an admirer of his endless curiosity and intellectual drive. Taylor Moulton has also become an incredible friend with whom I can discuss anything, often broadening horizons for what is possible in use of simulations in social sciences. I also thank Simon Friis, Summer Jackson, Ethan Poskanzer, and Hagay Volvovsky, who have sat through multiple presentations of my papers only to give incredibly thoughtful feedback. Emily Truelove and Jenna Myers have also been those to whom I can talk whenever I run into an impasse. I am also grateful to have Soomi Kim as a friend with whom I can talk about anything, as she has always been willing to lend an ear whenever I needed.

I also cannot express enough appreciation for the students who are more senior and have "paid forward" so much for me before leaving Sloan as fellow graduate students. When I first met Phech Colatat, he was in the middle of his job market season, but he took out hours to help me learn about the Sloan community. I did not know what that meant back then, only to realize recently what an incredibly generous gesture that was. He has been the same way since then even after he has left E62, and I can only hope I can be as half generous as him. Aruna Ranganathan is one of the people who convinced me to come to Sloan, and she has since been the unwavering example of what kind of researcher should come out of Sloan. Tristan Botelho and Josh Krieger have become like big brothers for me during the time we spent together at Sloan, and I am lucky to have them as those with whom I can have frank conversations about shortcomings of my research as well as exciting futures. I also thank Gokce Basbug for our friendship, from which I could find a lot of comfort that cannot easily be found. Mabel Abraham, Julia DiBenigno, Santiago Campero, and Abhishek Nagaraj have also been incredibly supportive along the way, giving me sage advice as well as being exemplary scholars themselves.

I have also benefited tremendously from help and support of people outside Sloan. In particular, I never knew that Daniel DellaPosta and I would be writing a paper together when we first met in the hallway of a dorm at University of Chicago. His support for my intellectual endeavor has been incredibly generous, as most of my papers have his name in the

acknowledgements section. Perhaps more importantly, working with him has made our friendship even closer; and I look forward to continuing our intellectual collaboration as well as friendship in many years to come. As a graduate from Sloan, Roman Galperin has always shown me ropes whenever I needed. Elena Obukhova has also been supportive throughout my intellectual journey during graduate school. Gino Cattani's encouragement along the way has been invaluable since he visited Sloan. Amanda Sharkey and Gary Herrigel have also been relentless advocates for my academic journey since my time in Chicago. I also express appreciation to all the people whom I met during the job market, who did not have any reasons to help and support me as they did but nevertheless did, among whom are Freda Lynn, Brayden King, Derek Harmon, and Frank Dobbin. Nathan Wilmers provided me with tremendous intellectual and moral support along the job market process as well, and I hope to continue our friendship down the road.

I am also lucky to have been able to spend years in Boston with friends from college, such as Bella Wu, Taweewat "Champ" Somboonpanyakul, Theresa Hwang, Kat Cheng, and Charis Crofton. Sinn Won Han and Kwan Woo Kim at Harvard Sociology have also been tremendously helpful and patient friends during my time in Boston, and our friendship will certainly continue wherever we are. I also appreciate Hannah Birnbaum and Daphne Chen for their support and care during good and bad times of my five years here. Lastly, I thank the "Boulder family" and especially Yumi Kim and Stella Kim for providing immense moral support ever since my time in Colorado. Their selfless and constant care for me is what keeps me happier during happy moments and afloat during difficult ones.

My acknowledgements would also not be complete without mentioning support staff at Sloan, some of whom had to deal with my nervousness during the job market. Gilly Parker has seen it all, from my last-minute appointment with Ezra to my panicking about applications, and she deserves credit for much of my sanity in the last three years. Hillary Ross, Davin Schnappauf, Helen Yap, and Patty Charest have also been incredibly helpful and understanding throughout everything, doing much more for students than we probably even realize and appreciate.

The most undeserved luck in my life is the family I was born into. While I always was (and still am) a brat in the family, my sister has always been the person to whom I could turn when problems arise as well as happy occasions occur. I am incredibly fortunate to have spent my life with my sister also in graduate school, and one of the fondest memories of my time here will be the countless phone calls that we have exchanged during my walk home from work. And finally, my deepest gratitude goes to my parents, Sung-Hie Koh and Young-Hwan Kim. When I first told them that I wanted to come to the United States for high school, they were vehemently opposed to the idea. But in the end, they supported me to come a half way around the globe, away from them, solely because they thought that that was best for me. I remember being stupefied by selflessness of my parents when I heard that they spent many wide eyed nights, worrying about me, upon my moving to Colorado. It tore their hearts to have their son thousands of miles away from them; yet they have endured everything just for my happiness. Their respect

and support for my pursuit regardless of their own comfort is an epitome of selflessness, and I would not have been able to achieve any – *any* – of what I have done without their being. As the smallest token of my gratitude to them, I dedicate this dissertation to my parents.


Minjae Kim
April 19, 2018
Cambridge, Massachusetts

# Table of Contents

*Chapter 1.* A Man Is Known by His Cup: Signaling Commitment via Insincere Conformity

*Chapter 2.* Faking It Is Hard to Do: Entrepreneurial Norm Enforcement and Suspicions of Deviance (coauthored with Ezra W. Zuckerman Sivan)

# Chapter 1

A Man Is Known by His Cup: Signaling Commitment via Insincere Conformity

> Interviewer: Do you drink and, if so, how often do you drink in after-hour gatherings?
> Respondent: We drink. We always drink. It's not mandatory that we drink, but we drink a lot, some more than others. Our company has this campaign that discourages us from drinking more than one type of drink at one venue and beyond 9 o'clock, but it is very rare that we stick to it. People almost always drink more and till later.
> Interviewer: Would you drink with your business partners even if you don't like to?
> Respondent: I wouldn't say that drinking happens because we like drinking…I doubt that we drink because any of us likes drinking.
> -Excerpt from author's interview with an employee of a major electronics company in South Korea, July 2016

## INTRODUCTION

One of the foundational notions of sociology is that social norms are an important driver of individual action and social order (Parsons 1937; Homans 1950; Gouldner 1960; Coleman 1990).[1] But while extensive research has documented the pervasiveness of normative conformity and its effect on valued outcomes, it is not always clear *why* individuals conform in the first place. Understanding why individuals conform to a particular behavioral standard (and why a group considers deviation to be problematic) is important because some norms persist even though they do not seem to provide an optimal way of coordinating actions (cf., Arrow 1971; Coleman 1990). One potential reason for such suboptimal conformity is individuals endorse the values the norms represent (Bernheim 1994; cf., Posner 2000, 2002). But even if a group and an individual do not endorse those values and would prefer to deviate from existing norms, there often seems to be a tendency to conform (Elster 1989; Hechter and Opp 2001:xvi; Bowles and

---

[1] I here focus on the definition of norms as behavioral standards demanded by others – "injunctive norm" – and not empirical regularity – "descriptive norm" (Cialdini 2003).

Gintis 2002). Why might people voluntarily conform to such "unpopular norms," i.e., norms from which most individuals in the group prefer to deviate (Bicchieri and Fukui 1999)?

This puzzling conformity to unpopular norms has been addressed partly by existing approaches seeing it as a misperception problem. For instance, Prentice and Miller (1993, 1996) suggest that individual college students misconstrue the binge drinking norm as widely endorsed and "falsely" conform so as to hide their private disagreement (also, see Asch 1956; Miller and McFarland 1987, 1991; Kuran 1995). This notion of "pluralistic ignorance" is thus sustained by a particular form of signaling: By creating the impression that one also endorses the values or preferences represented by a group's norms (even enforcing norms one does not endorse; Willer, Kuwabara, and Macy 2009), individuals signal that they are committed to the group (Posner 2000, 2002; also, Axelrod 1986; Bernheim 1994). The larger implication then is that unpopular norms persist *only when* they are *invisibly* unpopular norms, i.e., norms that are not known to lack (private) endorsement. Accordingly, Miller and Pretince (2016:340) conjecture that "providing people with [accurate] information …may, in turn, lead them to act differently." For example, Munsch, Ridgeway, and Williams (2014:57) argue that for some organizational norms that make flexible workplace arrangement more difficult, the "stigma [coming from adopting flexible work] may be attenuated by publicizing that many people only disapprove of flexible work because they, inaccurately, believe others do" (also, see Nunn and Sanchez de la Sierra 2017).

Thus, there are good reasons to expect conformity when a norm appears to represent a group's values or at least when it is unknown to be at odds with these values. By contrast, existing theory cannot account for conformity to norms that are *known* to be at odds with a group's values or group members' preferences. As an example of such a "visibly unpopular"

norm, consider the norm pertaining to drinking in after-hour gatherings in South Korea, as illustrated in the epigraph. The businessman does not believe that any of his colleagues like or endorse drinking; and if it is common knowledge (Chwe 2001) that the norm is unpopular, conformity with the norm should cease. But it does not: as described below, conformity persists and nonconformists continue to be punished. And this case is not unique: other examples of visibly unpopular norms include "ideal-worker" norms in organizations (Perlow 1997, 1998; Weeks 2004); norms around family structure in contemporary India (Derne 1992); and norms in authoritarian regimes (Wedeen 1999; Elster 1996). But if it is common knowledge that these norms are unpopular, why might people conform to them?

The main objective of this paper is to develop a theory that addresses this question. The key idea is that conformity to a visibly unpopular norm is an especially potent signal of commitment to interactants: who is more committed to a group than someone who is willing to override her own preferences to abide by its standards? Such signals are salient in contexts where actors are motivated to build embedded relationships in search of mutual gains (e.g., Uzzi 1997) but at the same time actors may fail to do so because they cannot be sure that their relationship partners are willing to accept (short-term) private losses for (long-term) collective gains (Hirschman 1970; Coser 1974; Azoulay, Repenning, and Zuckerman 2010). Potential relationship partners therefore confront the challenge of signaling their commitment to the collective gains. Conformity to a visibly unpopular norm provides a solution to this problem. No one thinks this conformity is sincere, but that is precisely the point (Spence 1973, 1981; Akerlof 1976; cf., Willer 2009): insincere conformity signals greater commitment to the collective interest than does nonconformity.

I elaborate on this theory in the next section. The main focus is on developing the logic of when and why signals of interpersonal commitment are important and hence when conformity to a visibly unpopular norm persists as a symbolic gesture. I also identify when conformity to a visibly unpopular norm might not be as useful: insofar as it serves as an indirect signal, it should be less valuable for those whose interpersonal commitment (or lack thereof) has already been more directly shown. After developing the theoretical framework, I analyze the norm around drinking in after-hour business gatherings in South Korea. Using qualitative evidence, I first show the "actor-side" of the mechanism, i.e., that Korean businesspeople are keenly aware of the signaling value of insincere conformity and that they strategically conform to reap such benefits. Next, using experimental evidence, I test the "audience-side" of the mechanism, i.e., that a conformist to the drinking norm appears as a more appealing partner for a business relationship when conformity to the visibly unpopular norm is explicitly or implicitly viewed as insincere, but not when it is viewed as sincere. Experimental evidence further shows that conformity to a visibly unpopular norm is not as useful when one's commitment has already been demonstrated directly. I conclude by noting macro implications of this paper's theory and findings. Most notably, this paper's analysis indicates that norms that are known to contradict the group's values or preferences might persist *precisely because of* its known contradiction and not despite that.

## THEORY

### Deepening the Puzzle

I begin by clarifying whether it is even possible for a norm to be visibly unpopular. One reason to be skeptical is that it would seem clear that if a norm were to sharply contradict the values or preferences held by group members, they can be expected to reject that norm.

Accordingly, one might be tempted to dismiss reports that the majority in a group prefers deviating from a norm with which they conform. The premise of the challenge here is that actions speak louder than words: if people continue to conform, they must believe in the norm.

It is noteworthy that this challenge is akin to the doctrine of "revealed preference" applied at the collective level. This doctrine essentially implies that whatever human beings do is what they prefer to do, and it thus assumes away the possibility that choices may be highly constrained: As Sen ([1977]1979:323) describes the premise of revealed preferences, "[t]he rationale of this approach seems to be based on the idea that the only way of understanding a person's real preference is to examine his actual choices, and there is no choice-independent way of understanding someone's attitude toward alternatives." But many contexts impose constraints on choices; thus, one may choose the better of two bad options even when one would prefer to a third option that one cannot obtain given the constraints.

In fact, it seems common that a group sticks with a coordination standard on which it has converged, even if it *knows* of a better standard. One example is the well-known case of QWERTY keyboard. Once the default arrangement of computer keyboard (QWERTY) was adopted, switching to the more efficient keyboard (e.g., DVORAK) has been difficult due to the very high cost of retraining and retooling (David 1985). The key idea is that individuals continue coordinating on the "suboptimal" standard (i.e., behavioral standard from which they would prefer to deviate) even though they recognize a good alternative but cannot navigate to that alternative given the social costs of doing so. The applicability of this logic to social norms is straightforward and is most salient when members of a group or community are confronted with evidence of another group or community that has succeeded in coordinating around a more attractive norm for one of its core activities. When the costs of shifting off the existing standard

are not too high, it is possible for the alternative norm to be adopted (Bendor and Swistak 2001). But in many cases, groups seem to be stuck in such suboptimal standards even though individuals prefer to deviate (cf., Schelling 1981).

Examples of group norms that provide such "visibly suboptimal" coordination standards (i.e., coordination standards that are known to be suboptimal) include "ideal-worker" norms such as facetime and working overtime. Even though one explanation for persistence of these norms is the (cultural) consensus that "ideal workers" in organizations endorse the norms' values (Blair-Loy 2003; Turco 2010; Williams, Berdahl, and Vandello 2016), it is commonly found that some of these norms are already known to be at odds with the values endorsed by group members. In fact, organizations often try to *replace* them with alternative work arrangements because individuals (including supervisors) collectively disapprove of such norms and see them to be unproductive. For instance, in describing norms in the culture of a British bank, Weeks (2004:2) reports: "Never once during the fieldwork that I conducted in [the Bank] did I hear it mentioned in a positive context…even though the Bank has spent large sums on repeated (and sometimes overlapping) programs of culture change…'the Bank hasn't really changed in three hundred years.'" Similarly, as describing the norm of working overtime, Perlow (1998:343) reports in her ethnographic account of a white-collar organization: "all of the engineers felt there was no practical reason for their project team leader to have stayed…[i]n fact…his presence might have slowed the team's progress." Other accounts of ideal-worker norms also indicate that people often see these norms to be at odds with their organization's values and members' preferences (Lawler 1986; Bailyn 1993; Perlow 1997, 1998; Kellogg 2009). But conformity seems to persist even then.

Similarly, individuals often recognize that others in the group prefer to deviate from norms in exclusive groups such as Greek organizations on college campuses, activist organizations, and gangs; nevertheless, conformity often persists. Although an existing explanation based on research on cognitive dissonance (Festinger 1957; Aaronson and Mills 1959; Deci 1971) is that norms such as hazing rituals signal a group's endorsement of the values underling a norm; and that coming to endorse such values is part of the group socialization process (Van Maanen and Schein 1979; Cialdini 1984; Cialdini and Trost 1998), these norms often clearly contradict the group's implicit or explicit preferences. For instance, when the Korean peninsula was occupied by Japan in the early 20$^{th}$ century, cutting off part of one's ring finger persisted as a norm in a group of twelve Korean activists called *"Dongeui Danjihwe"* (translated as "association of uniform willingness and cut fingers").[2] It seems incredibly difficult to believe that members of the group enjoyed such physical pain in the first place; or that those activists were socialized to enjoy such physical pain. Yet, even though such a norm seems to continue to lack the group's endorsement, conformity to such a norm, along with other norms in gangs (Gambetta 2011), often persists.

Note that in both cases of ideal-worker norms and norms in exclusive groups, those higher in the hierarchy *also* disapprove of the norms and see them as unproductive for their own (and the group's) good but continue to reward those who conform. That is, they ostensibly want the norms to disappear, but their actions reinforce persistent existence of the norm. Thus, a good explanation for persistence of visibly unpopular norms should be able to account for why people

---

[2] Bearing a striking resemblance, the Talmud illustrates a story that young recruits of Simon bar Kokhba – a Jewish leader of the Bar Kokhba movement against the Roman Empire in 132 CE – cut off one of their fingers to join the group.

higher in the hierarchy also help perpetuate norms that they themselves despise. Other examples of such visibly unpopular norms include:

*Norm around family structure in contemporary India*: Derne (1992) documents that even though Indian men resent and prefer to deviate from the norm of building a "traditional family"—a norm they see as both dysfunctional and unpopular, conformity persists.

*Norms in totalitarian regimes*: Wedeen (1999) reports that even though Syrian citizens recognized the regime to glaringly contradict values and preferences of the constituents and even its own political establishment, they still went through the motions and showed their support for the regime by conforming to the norms (also, see Elster 1996; cf., Kuran 1995).

*Footbinding norm in China*: Mackie (1996) describes how most Chinese did not endorse footbinding as a practice and recognized that others also did not endorse it; yet conformity to the norm persisted for many years.

In short, this paper's puzzle can be restated as: Why is it so hard to abandon suboptimal normative standards that individuals collectively despise? That is, why do group members reward conformity when they themselves would prefer to abandon such coordination standards? The challenge here is to develop a theory that addresses why group members might actively reward those who conform even when their own preferences for an alternative are common knowledge.

## Signaling Interpersonal Commitment via Insincere Conformity

To make progress in resolving this puzzle, I extend signaling theory to account for when and why such conformity might occur; and it starts from recognizing when actors have a reason to signal their commitment to one another but not to the values that the norm ostensibly

represents. One context in which interpersonal commitment is salient is when individuals attempt to build embedded relationships within and across organizations (Granovetter 1985; Powell 1990). In such situations, actors face the challenge of convincing others that they will stay committed to a relationship (Hirschman 1970; Coser 1974; Turco 2010; Azoulay et al. 2010; Phillips, Turco, and Zuckerman 2013). In particular, actors often need to credibly signal to one another that they are the type of relationship partner who will serve the larger interest over their own private interest, as mutual gains from embedded relationships are often possible only when actors are willing to hold off or even sacrifice their short-term private gains (Uzzi 1997, 1999). Yet mere assertions of interpersonal commitment cannot be taken at face value—they could be "cheap talk" (cf., Crawford and Sobel 1982; Kreps and Wilson 1982; Frank 1988). Therefore, a prevalent challenge in our social and economic lives is how to credibly signal one's commitment to the collective interest.

I argue that conformity to a visibly unpopular norm can serve as such a credible signal. Unless there is more direct information about a conformist's preferences, conformity to a visibly unpopular norm can be assumed to be "insincere" because it is common knowledge that most group members do not endorse the norm. Such insincere conformity is thus a form of sacrifice to group demands over personal preference. Insincere conformity to a visibly unpopular norm then signals not that one endorses the norm's values but that one is committed to the demand from one's interactants. Consequently, insofar as this signaling value is important for actors, they can be expected to undertake it.

This logic can be clarified by considering a situation where an actor appears to sincerely *endorse* a visibly unpopular norm. In such a situation, two different motives might be inferred: (a) she conforms because one enjoys doing so; and/or (b) she conforms because she is willing to

conform to the behavioral standards demanded by the collective. Since motives in such a situation are ambiguous, conformity does not effectively signal commitment to the collective. By contrast, the inference (a) is not available for "typical" conformity to a visibly unpopular norm, since it is assumed to be "insincere." Not only does inference (b) then become the likely candidate, but one's commitment to the collective interest is highlighted even more: costliness of conformity suggests that one's commitment to serve the demand from the collective *overrides* one's regard of one's own preference (Spence 1973, 1981; Akerlof 1976; Landers, Rebitzer, and Taylor 1996; Henrich 2009; Willer 2009).[3] That is, conformity to a visibly unpopular norm persists because such presumed-to-be-insincere conformists appear more committed to collective interests than nonconformists (i.e., those who do not seem willing to serve the demand from the collective).

To be sure, conformity to a visibly unpopular norm is an indirect way of signaling one's commitment to the collective interest. In contexts where more direct indications may be available, conformity to a visibly unpopular norm as a signal might not be as useful.[4] Similarly, such conformity is less valuable when one does not want or need to show commitment. Therefore, the need for conformity to a visibly unpopular norm might be salient only in situations where actors have not directly shown their commitment to the collective interest but

---

[3] While these studies on costly signals usefully articulate why costliness of conformity enhances credibility of commitment, they are not able to account for conformity that is *not* intended to signal shared endorsement of the group's values or preferences: "numerous authors...have suggested that costly religious acts can be seen as honestly conveying information about the religious signaler's commitment to the religious and moral precepts of the community" (Power 2017:83; also, see Rappaport 1999). Part of this paper's contribution thus is to identify visibly unpopular norms as contexts in which costly conformity is used to signal interpersonal commitment and not commitment to the group's values or preferences. I further discuss this point in the General Discussion section below.

[4] This however is not to say that the need for insincere conformity to a visibly unpopular disappears completely once more direct indications are available, since commitment may very rarely be demonstrated without doubts (Vaughan 1986). The diminishing value of insincere conformity, therefore, should be understood as a matter of degree. This is more directly tested in the empirics below.

need to do so.[5] This logic is summarized in Table 1. In sum, this paper's theory can be expressed by the following proposition:

> *Proposition: Conformity to a visibly unpopular norm will be assumed to be insincere and it thus signals greater commitment to the collective interest than does nonconformity. Insofar as one recognizes and seeks this signaling value, one will conform.*

Table 1 around here

A key implication of the proposed theory is that misperception of popularity might be sufficient but not necessary for the persistence of an unpopular norm. Recall the implication of the notion of pluralistic ignorance: once there is no longer a misperception problem (i.e., once individuals recognize a norm's contradiction to the group's values), conformity will stop. By contrast, the proposed theory implies that common knowledge (Chwe 2001) of a norm's unpopularity provides a strategic opportunity to signal interpersonal commitment (cf., Swidler 1986). That is, the conditions under which conformity should cease according to the theory of pluralistic ignorance are precisely the conditions that propel further conformity and trap groups in suboptimal coordination standards.

## EMPIRICAL SETTING AND RESULTS

The overarching empirical goal here is to see whether actors would not conform to the visibly unpopular norm because they do not need to feign their endorsement (as suggested by

---

[5] This implication runs counter to a prominent line of research in social psychology – i.e., investment model – which is an extension of social exchange theory (Homans 1961; Blau 1964) and empirically tested by Rusbult (1980, 1983). This argument predicts that individuals will feel attached to the relationship proportionally to the investment they put in, implying positive correlation between investment and attachment to the relationship. I therefore use the evidence below to adjudicate these arguments and rule in that insincere conformity to a visibly unpopular norm is more frequent and/or intense in nascent relationships.

extant literature) or whether actors would conform to the visibly unpopular norm because they want to signal their interpersonal commitment through showing personal sacrifices (as suggested by my theory). I take on this empirical task by examining an ideal-worker norm that is also prevalent in different contexts – the norm prescribing excessive drinking in after-hour gatherings in South Korea (Korea hereafter). This is a suitable "strategic research site" (Merton 1987) for several reasons. First, the drinking norm is publicly recognized as widely unpopular both among workers and organizations in this context. For instance, in a study conducted by the Korean Ministry of Employment and Labor using a representative national sample of the working population in September 2013, when asked "what do you want to avoid the most during your after-hour gatherings?", 69.1% of the sample (n=3,302) responded "after-hour gatherings that are centered around drinking" ("Work net online survey").[6] More importantly, not only is there private disagreement, but there also is common knowledge on how drinking in after-hour gatherings is undesirable, as often described in the media (e.g., Cha 2015) and shown in Study 1 below. I further show that not only do individuals but organizations also actively try to discourage conformity to the drinking norm.

Second, despite its visible unpopularity, excessive drinking in after-hour gatherings is an omnipresent feature of businesspeople's lives in Korea. In the same 2013 survey that reported unpopularity of the drinking norm, 67.3% of the sample[7] responded that they are still participating in "after-hour gatherings for which drinking is the central activity." Similarly,

---

[6] Other choices were "picnic-style gatherings in suburb" (13.4%), "gatherings that involve recreational sports" (6.3%), and "foodie-tour gathering" (5.5%), among others. The question allowed respondents to pick only one out of many options as something that they want to avoid in after-hour gatherings. Since it is highly possible that they want to avoid excessive drinking but it is not the *most* undesired activity in after-hour gatherings, 69.1% is likely a conservative estimate of the Korean businesspeople's distaste for excessive drinking in after-hour gatherings.

[7] This answer is again likely a conservative estimate of prevalence of drinking in after-hour gatherings, as shown in interview data below, since drinking can occur without being the central activity of gatherings.

Korean popular media such as documentaries and soap operas describing Korean businesses almost always portray excessive drinking in after-hour gatherings as an important activity of one's work life. Western media have taken a note of its omnipresence as well (e.g., Cha 2015). For instance, a popular advice book for Westerners called *The Korean Way in Business* describes that "[drinking in after-hour gatherings] became an integral part of developing and maintaining interpersonal relationships between coworkers, suppliers, and customers…[it] continues to play a key role in the etiquette and ethics of South Korean businesspeople and government officials" (De Mente 2014: 157-158). Consequently, drinking in Korean after-hour gatherings serves as a theoretically relevant and practically important phenomenon in studying conformity to visibly unpopular norms.

Note that these prominent features of the setting are not to suggest that *all* attendees of after-hour gatherings consider drinking in such a setting as an undesired activity; or that *all* attendees drink. Several businesspeople interviewed for Study 1 below said that it is quite possible that one drinks in after-hour gatherings because one wants to. Also, they described conformity to the drinking norm as voluntary, so it is not surprising that there are nonconformists. However, it is worth highlighting that all of my informants said that they have drunk in after-hour gatherings, *even when their distastes for drinking were known by others and they knew that others disliked drinking*, as I will show further below in Study 1. It is also important to clarify that even though many of them might like drinking in other settings (e.g., dinner with friends or colleagues), they expressed severe discontent with the norm, for it often takes place 10+ hours of work, feels like extended work, and requires them to drink more than they want to. My sample is not meant to be a representative one (see below for sample characteristics); yet this paper's question and theory address why and when removing the

misperception problem is *not sufficient* to stop conformity. Therefore, my analysis focuses on why actors conform to the drinking norm *even when* they recognize it as unpopular, even if they might in rare times see it as popular.

It is also worth noting how this norm is applicable to wider contexts in two different ways. First, this norm of excessive drinking in after-hour gatherings is prevalent in Western contexts as well, especially in industries such as consulting, law, banking, and accounting where clientele relationships are forged (Fincham 1999; Werr and Styhre 2002). At the same time, the norm is widely recognized as unpopular, making it a visibly unpopular norm: as described in the recent *New Yorker* article, "After-hour drinks, second only to working breakfasts, are one of the great nuisances of office life – particularly when they're planned, and in fields where they're quasi-mandatory" (Collins 2016). Second, this norm prescribing excessive drinking is analogous to other ideal-worker norms such as working overtime and facetime. Conformity to them often persist, *even when* they seem to have no clear functional value and contradict the group's preferences (e.g., Lawler 1986; Bailyn 1993; Perlow 1997, 1998). My empirical investigation is meant to reflect such general characteristics of visibly unpopular norms more broadly.

Lastly, while one might suggest that the need for embedded relationships is especially intense in a Korean context (or other East Asian countries; e.g., Xiao and Tsui 2007), such a need seems present in Western contexts as well (e.g., Podolny and Baron 1997; Azoulay et al. 2010). In particular, Yamagishi, Cook, and Watabe (1998) show via experimental evidence that their Japanese subjects perceive their institutional setting to be uncertain and thus sought embedded relationships more than their American equivalents; but when people of both countries are in an equally uncertain context, they equally pursue forming embedded relationships (also, see Kollock 1994; cf., Yamagishi and Yamagishi 1994). Consequently, while unlikely a

conservative setting, the norm around excessive drinking in Korean after-hour gatherings might still represent a general case of conformity to visibly unpopular norms.

Two complementary studies validate the proposition in this setting. Study 1 presents qualitative evidence from interviews on the "actor side" of the mechanism – i.e., whether the norm around excessive drinking in after-hour gatherings is indeed visibly unpopular and whether Korean businesspeople recognize the signaling value of conformity. It also examines whether patterns of conformity (and nonconformity) reflect the proposition. Study 1 thus establishes the "existence proof" that actors in the setting very much have strategic motives for insincere conformity to the visibly unpopular drinking norm and try to reap practical benefits from it (e.g., favorable career outcomes; success of interorganizational relationships). Afterward, Study 2 presents causal evidence from survey vignette experiments on the "audience side" of the mechanism. While interviews in Study 1 help get at whether and why conformity to the drinking norm in after-hour gatherings might persist, it might be difficult to directly test that it is specifically *insincere* conformity to visibly unpopular norms (and not any conformity) that provides its strategic value. Therefore, Study 2 uses online experiments to directly manipulate the perceived preference for drinking in after-hour gatherings (i.e., how much one likes or dislikes drinking in such a setting). Additionally, there needs to be a more direct causal test to validate that the effect of insincere conformity to visibly unpopular norms is *sufficiently* driven by the perception of interpersonal commitment, while other mechanisms (e.g., emotional intimacy; physiological effect of drinking) can certainly be complementary. Study 2 therefore provides a controlled experiment and isolates the causal effect.

**Study 1: Interview**

*Study 1 Data and Methods*

For Study 1, I conducted 42 interviews between May and November of 2016. While all informants had previously worked in Korea and most informants were working in Korea at the time of interview, eight of them were working outside of Korea, getting additional training (e.g., MBA), or had retired. All interviews were conducted by the author, most of them by phone (n=38) and a few in person (n=4). Interviews lasted from thirty minutes to two hours. The average duration was a little over an hour.

For sampling, I first created categories that I expected to have variations on the need to signal interpersonal commitment and sampled on these dimensions (Trost 1986). The first dimension is the informants' position in the hierarchy. I expected those lower in the hierarchy to have a greater need to signal their commitment, since their commitment is likely evaluated by those higher in the hierarchy, whereas those higher in the hierarchy might be rarely evaluated by their subordinates. Also, by varying where informants are placed in the organizational hierarchy, I examined whether informants across the hierarchical order have similar perceptions of what drinking in after-hour gatherings means. By doing so, I further examined that the conformity is not driven by the very existence of hierarchy but the way in which evaluation happens via hierarchy. Another dimension is whether or not one faces clients. I expected those in client-facing roles to face this challenge more often and intensely, since those in interorganizational relationships are likely to have fewer organizational or structural tools to assure each other's interpersonal commitment. Lastly, the other dimension is the organizational type, since I wanted to make sure that this logic of actions and inferences was not just present in one organizational

type but all. Therefore, I sought variation on this dimension. Sample characteristics are presented in Table 2.

Table 2 around here

Based on these categories, I first used personal contacts to recruit informants and then used a snowballing strategy to complete sampling. There was no particular reason to suspect that the sample likes or dislikes drinking more than the average working population in Korea; but because I was concerned that my sampling strategy would give me a sample with homogeneous tastes for drinking by oversampling on individual networks, I limited the number of referrals from each contact to three. Everyone but one whom I contacted responded to my request for an interview. Two of them asked to be interviewed using an email questionnaire (in lieu of phone or face-to-face interview), so while their responses were largely consistent with responses from phone or in-person interviews, I excluded them from the analysis. I stopped adding to my sample when additional interviews were no longer contributing new insights and I believed I had reached theoretical saturation (Glaser and Strauss 1967).

In requesting interviews, I sent an email asking for time to discuss their corporate culture and social lives at work. I specifically chose not to disclose drinking in after-hour gatherings as the topic of the interview. I did this mainly because I did not want informants to think about ways to rationalize why they drink in their after-hour gatherings, despite their distastes. Similarly, during interviews, I asked them to describe after-hour gatherings at their work without specifically mentioning drinking. It is worth noting that while I brought up drinking only when informants mentioned it first, it was mentioned in *all* interviews as a primary feature of their after-hour gatherings at work. Once the topic was brought up, I made sure not to directly ask *why* they drank, since doing so might have forced them to rationalize their conformity. I also asked

them for their "third-order" inference – what informants think others would think – rather than asking what they would think of those who drink or do not drink in a certain situation (Ridgeway and Correll 2006; Correll et al. 2017). By doing so, I tried to help informants feel comfortable giving explanations consistent with their true beliefs and not ones they thought they were supposed to give. Also, asking third-order questions was important because the interview questions addressed what my informants perceived to be normative in the community and not necessarily what they preferred to do or what they themselves liked about others.

Lastly, while about half of informants were women, I focus in the analysis on the general (and not gender-specific) effect of conformity to the norm around excessive drinking. That is not to say that women are on equal footing with men when it comes to excessive drinking: in the data gathered through interviews, there were gendered differences in (i) the degree to which excessive drinking was perceived to violate one's preference; and (ii) the degree to which they were given the chance to conform to the norm, as I show below. Nevertheless, female informants reported a similar rationale for why they conform to the norm around excessive drinking and similar patterns of conformity. Therefore, I focus on the general effect of insincere conformity in the analyses and investigate the gender-specific issue further in the Study 1 Discussion below. All interviews were conducted and transcribed in Korean, and I analyzed the transcripts using the qualitative software Atlas.ti. All quotes shown below were translated to English by the author.

*Study 1 Results and Analysis*

*Drinking as an Omnipresent Feature of Korean After-hour Gatherings*

It is first important to situate what Korean after-hour gatherings look like. Korean businesspeople reported a wide range of frequency at which after-hour gatherings happen, from

every other day to once a month. The number of attendees reportedly varies widely as well, from

five to sixty, and attendees are usually members of the same team or external business partners

(latter if in an interorganizational after-hour gathering). Almost always, all expenses are paid for

by their firms, although there are often limits imposed by the organizations on how much can be

spent for each after-hour gathering. Those in client-facing roles said that they almost always go

to after-hour gatherings with their clients in the beginning, during, and at the end of a project.

Although these gatherings are often put on the official schedule of their department or firm one

or two weeks in advance, many informants said that many gatherings in their workplaces also

happen impromptu. Either way, however, all informants emphasized that they are expected to

attend gatherings at most times, especially when an important evaluator is present (e.g., someone

high in the hierarchy; important client). One retiree from a major chemical engineering company

recalled:

> I want to go home, after working on Saturday night. And then, someone says that we should get
> drinks…If I say that "I don't want to go, so I'll go home," then I can feel their scorching looks on
> my back. What they are going to say after I leave for missing the gathering, isn't that obvious?

*All* informants said that once everyone arrives at a gathering (most often at a restaurant

for dinner), drinking is almost always to follow. For instance, a junior-level software engineer

said that "After-hour gatherings equal drinking in Korea. The only thing that varies is how much

you drink." These responses are in fact not surprising to those who are familiar with the Korean

business context, while the duration and intensity of drinking in gatherings seem to vary widely.

Most informants reported that they sometimes participate in multiple rounds, going from

restaurants to bars to karaoke bars, while at other times they leave after drinking only a little at a

restaurant. I exploit this variation to see when the more intense conformity happens.

It is also important to situate how one is usually put in a situation of drinking in this

setting. For those more familiar with the US setting, it is more intuitive to picture a cocktail party

to which one can attend without necessarily drinking. In such a setting, it might also be difficult

to know whether and how much other people are drinking. By contrast, attendees at Korean

after-hour gatherings are often in a situation where their drinking is directly observable. In

particular, attendees at gatherings often exchange drinks with one another, in a practice called

"circulating a glass." This often entails one's taking a shot (usually Korean rice wine called *soju*

or beer mixed with *soju*) and the other's using the same (or different) shot glass to take a shot

right afterward (or simultaneously). This practice is also often done with more than two people,

in which case everyone around the table takes a shot consecutively or concurrently. This practice

makes one's conformity to drinking (or a lack thereof) highly observable. For instance, in a

conversation with a junior-level statistician at a major bank, he said:

> When we begin our gatherings, we make and circulate the soju bomb with one another. And then, we say cheers. After that, I give my empty glass to the head of the department, and then he makes one for me. After that, another person makes one. And we keep drinking like this.

In part because of close monitoring enabled by the practice of "circulating a glass,"

attendees at a gathering almost always drink, often excessively (i.e., more than they like).

Despite its oddity to outsiders, however, most informants expressed no surprise that attendees of

a gathering almost always drink. For instance, in a conversation with a mid-level employee in a

major chemical company:

> Interviewer: What percentage of people would you say drink in after-hour gatherings?
> Respondent: I'd say everyone. Yeah, everyone.
> Interviewer: Does really every single person drink? What about exceptions?
> Respondent: [Exceptions are] really rare. Everyone drinks, except for those who don't drink for religious reasons or those who physically cannot drink. In those cases, obviously no one encourages them to drink, since [not drinking] is inevitable. But people still like those who drink.
> Interviewer: Then how do you express the desire not to drink, when you don't want to drink?
> Respondent: If someone gives one a shot, it is rare that one rejects it…Even if one does not feel well that day, one at least receives the shot glass.

Lastly, even though there is pressure for attending after-hour gatherings and drinking, all

informants described their drinking as voluntary. For instance, many reported that others often

"suggest" they drink, or "inquire" whether they would drink, and once they refuse to do so, few

follow up *requiring* them to drink. For instance, one mid-level manager in a pharmaceutical

company said: "When one says 'I don't want to drink,' no one would say 'you must drink.'"

Another said "At the end of the day, it is you who decided to drink or not." These descriptions on

the voluntary nature of drinking are especially helpful in establishing this setting as a relevant

case.

*Drinking in After-hour Gatherings as a Visibly Unpopular Norm*

Even though drinking in after-hour gatherings happens on a monthly, weekly, or even

daily basis, most of my informants suggested that they (i) dislike drinking in after-hour

gatherings, (ii) recognize that few others like drinking in those gatherings, and (iii) recognize that

others know that they themselves probably do not like drinking in such a setting. For instance, in

a conversation with a mid-level employee in a major electronics company's strategy department:

> Interviewer: What do you think people think of drinking in after-hour gatherings? Do you think
> people like it? Or do you think people dislike it?
> Respondent: I really dislike it, and most people probably dislike it too.

Another informant – a junior-level software engineer in a medical device company – said that

"Most people don't like it. Even increasingly more so nowadays." It is also important that not

only junior-level but also senior-level workers often expressed discontent with the norm. A

senior-level manager in a pharmaceutical company put it as: "Everyone including me dislikes

drinking in after-hour gatherings. Very few people like it."

For specific reasons why they do not like drinking in after-hour gatherings, informants

most often referred to the excessive amount of alcohol consumed in those settings, especially

after infamously long hours of work in Korean workplaces.[8] Many said that even though they

enjoy drinking with their friends, they most often want to avoid excessively drinking in after-

hour gatherings after working for 10+ hours of work. For instance, as one junior-level

accountant recalled the latest after-hour gatherings:

> Depending on which client it is, even partners [who are more experienced in drinking] come to
> work sick the day after [the after-hour gathering]. Because they drank way too much. They must
> have been exhausted.

Another informant – a junior-level employee in a governmental agency – provided a similar

account, saying: "I think that drinking in after-hour gatherings in and of itself might be okay, but

it's the excessive amount of drinking that people don't like. It really depends on how much you

can freely decide how much you can drink." These accounts then suggest that the discontent

comes from the fact that drinking in after-hour gatherings tends to be excessive (i.e., more than

one wants).

Other informants additionally said that their distaste for drinking is exacerbated because

drinking in after-hour gatherings takes place as part of their work obligations. For instance, in a

conversation with a mid-level manager in a governmental agency:

> Interviewer: What about drinking in after-hour gatherings do people not like?
> Respondent: It probably depends on who you are. Some people probably don't like drinking in
> itself, and others probably don't like it because they think that it's extension of their work.
> Especially since after-hour gatherings happen after work hours.

Similarly, a junior-level consultant said "When you drink with others at work, it just is not

comfortable. You have to think about what they think of you even when you are drinking that

much." These accounts then suggest that many attendees do not like drinking in those after-hour

gatherings because it is perceived not as extra work.

---

[8] Work hours in Korea are consistently rank high among non-developing countries, totaling 2113 hours
per year in 2015 (3rd among OECD countries). For reference, annual work hours for workers in the US
was 1790 hours in the same year (OECD Statistics 2017).

It is worth noting that the widely recognized and shared distaste for drinking in after-hour

gatherings is also highlighted by organizational policies. In particular, many firms have tried to

cut down on excessive drinking in after-hour gatherings by limiting the amount of money

allocated to those gatherings or having employees report when there is excessive drinking.

Alternatively, firms have adopted different policies to discourage their employees from drinking

excessively in after-hour gatherings. Most prominently, one of prominent *chaebols*[9] has adopted

a policy called "119" (the equivalent of 911 in Korea), which encourages its employees to have

only one type of drink, go to only one venue, and go home by nine o'clock (see Figure 1). This

campaign then shows that the organization at least ostensibly dissents and tries to stop

prescription of and conformity to the norm. To the organization's dismay, however, informants

working in the *chaebol* reported that few actually subscribe to the policy. For instance:

> Interviewer: Do you drink and, if so, how often do you drink in after-hour gatherings?
> Respondent: We drink. We always drink. It's not mandatory that we drink, but we drink a lot,
> some more than others. Our company has this campaign that discourages us from drinking more
> than one type of drink at one venue and beyond 9 o'clock, but it is very rare that we stick to it.
> People almost always drink more and till later.

Figure 1 around here

*Ruling out the Extant Account: Drinking Not as a Way of Feigning Endorsement*

Recall the argument from the notion of "pluralistic ignorance": individuals conform to

the norm because they misconstrue the norm as popular and conform so as to feign their

endorsement (Prentice and Miller 1993; cf., Posner 2000). However, my informants do not seem

---

[9] *Chaebols* are large business conglomerates that are typically owned by a family or a few. Firms that are most familiar to people outside of Korea (e.g., Samsung, Hyundai, LG) tend to be *chaebols*. Most often, *chaebols* and small and medium enterprises (SMEs) in Korea engage in different markets, even if they are in the same industry. For instance, Hyundai Motor Company assembles different parts of automobiles that are supplied by SMEs in their supply chain.

to conform in order to feign their endorsement; and they alluded that their drinking in after-hour

gatherings is not about hiding their distaste for drinking. For instance, a junior-level accountant

said that his "[clients] probably know that we don't like drinking." Similarly, in the conversation

with a mid-level employee in a major electronics company:

> Interviewer: Would you drink with your business partners even if you don't like to?
> Respondent: I wouldn't say that drinking happens because we like drinking... I wouldn't say that drinking happens because we like drinking...I doubt that we drink because any of us likes drinking.

These accounts then show that businesspeople in Korea infer "insincere conformity" from

others' drinking, based on their inference that drinking excessively in this setting is widely

dissented. This inference is further confirmed in Study 2 as well.


*Then, Why Drink? Demonstrating Interpersonal Commitment via Drinking in After-hour*

*Gatherings*

Instead, most of my informants suggested that drinking in after-hour gatherings is

important precisely because explicitly and publicly overriding their known distaste for drinking

is important. For instance, a junior-level consultant said:

> Interviewer: What do you do when you don't want to [drink at after-hour gatherings]?
> Respondent: As I mentioned before, there are times when one doesn't want to drink because one is sick or something. Even then, people drink so that they can show their loyalty. It's like "I'm sick, but I respect you, so I drink despite feeling bad."

Similarly, a mid-level employee in a governmental financial agency said:

No one says "don't drink," even though they might say "drink only a little today." So, you pretty much need to drink even if you don't feel well. I think that you have to drink despite feeling bad, and only then people would think that, "ah, this person is really sincere in how he thinks of me."

A mid-level banker also said that via drinking, one can express loyalty precisely because others

know that one does not like drinking:

> I think that when colleagues circulate a glass with me, taking that glass and not rejecting that glass is about showing loyalty...it's about not saying "no" when someone suggests circulating a

> glass with me. And normally, in Korean style, one needs to bottom-up the drink. It's like, as they see me bottom-up the drink and cringe, they feel "ah, this person is really willing to obey me." Like, doing a bottom-up even though alcohol is bitter.

Consequently, these accounts suggest that drinking is perceived as an effective signal precisely because of the shared and widely recognized distaste for drinking excessively in such a setting.

It is important to note that many of my informants had *strategic* motives while making their decisions about drinking in this setting. For instance, a junior consultant expressed fear that his unwillingness to drink might be interpreted as a lack of loyalty:

> Interviewer: How do you respond when you don't want to drink or can't drink?
> Respondent: From my observation, if I can't drink because of some health reasons or work reasons…I can be excused. But if the reason was "I don't like drinking" or "I don't want to drink today," it's not perceived well. (Why not?)…it's because not drinking is perceived as lacking loyalty to them.

Moreover, in order to enhance its strategic value, many informants said that they try to make their distaste for drinking visible to some extent. By doing so, they could make their conformity appear more likely to be motivated by their commitment to their interactants (and not their preference for drinking).[10] For instance, a former junior-level employee at an IT company expressed frustration over how she had to drink in the beginning of her tenure to appear loyal to the group but at the same time how she then seemed to *enjoy* drinking and consequently lost drinking as the strategic tool to credibly signal her commitment:

> What I did regret in my two years [in my last workplace] is that…when I first interned, I felt like I needed to get along with my team when I was an intern, so I drank a lot, and they expected the same thing of me when I returned as a full-time employee. And, that's fine, in the beginning. But I guess that becomes your life, I had a reputation for [enjoying] drinking.

Lastly, it is worth highlighting that many informants were cognizant of the public aspect of their drinking behavior. By publicizing their conformity, they alluded that their commitment to the group can also be publicized. For the same reason, they thought it important not only to

---

[10] Nevertheless, there might be a limit in how far this performative local action can go. I further elaborate on this point in the General Discussion section below.

drink in front of their immediate interactants, but also to show other possible audiences that they are drinking with one another. For instance, a mid-level banker said:

> Interviewer: Do you think people usually like drinking in after-hour gatherings?
> Respondent: There are some who like drinking when they meet with friends, but there are few who like drinking in after-hour gatherings…But I think heads of department or teams want to affirm that their teams are united by seeing that their subordinates drink together…And I think by doing so, they want to see that when they are heads of the department, their departments are in unity.

*Variation: When Do They Conform More to the Drinking Norm? When less?*

These accounts validate that conformity to the visibly unpopular drinking norm is at least partly driven by their desire to signal commitment to their interactants (and not the desire to feign endorsement). Another way of ascertaining the value of drinking against one's preference is to look at conditions under which drinking actually happens more often. Examining the variation is especially important in this setting, since my informants might fake their distaste for drinking to me in order to appear more socially desirable: Even if they in fact enjoy drinking, they might not have said so during interviews with me. Alternatively, they might drink because they want to be more straightforward with their colleagues while being able to plausibly deny being unprofessional by pretending to be drunk (Van Maanen 1991). Therefore, I here present variations on when drinking in after-hour gatherings happens more frequently and which explanations can best account for such variations. By doing so, I do not have to solely rely on informants' expressed rationales in identifying motives for drinking against one's distaste (cf., Jerolmack and Khan 2014). Note that while this paper's proposition suggests that conformity would be more intense when the need to signal commitment is higher, I did not have specific predictions on exact empirical conditions; therefore, this part of the empirical analysis is more inductive in nature. The inductively emerged variations are visually represented in Figure 2.

Figure 2 around here

*i) When the relationship is nascent*

The most prominent dimension of variation that emerged from the interviews was

familiarity with one's relationship partner, within and between organizations. And this variation

is especially useful for the following theoretical reason: The investment model in social

psychology, formalized by Rusbult (1980, 1983), predicts that individuals will feel attached to

the relationship proportionally to the investment they put in, implying positive correlation

between investment and attachment to the relationship. By contrast, this paper's theory predicts

that drinking would happen less often in older relationships, *ceteris paribus*. And validating the

latter prediction, most of my informants said that it is especially important for someone new at

an organization to drink in after-hour gatherings, since his or her drinking is seen as an indicator

of what he or she is like as a colleague. For instance, in a conversation with a junior-level

consultant:

> Interviewer: Are there instances where it is more important to drink? Or less important to drink?
> Respondent: If you are a new hire, or if there are not enough data to know who you are, then if
> you say that you can't drink, then people would think "What? A new hire says that he can't
> drink?"...But independently from drinking, let's say that that you turned out to be a dependable
> co-worker. Then there are few negative consequences associated with not drinking, and you
> don't have to drink.

The same source of variation existed for drinking in after-hour gatherings of an

interorganizational relationship as well. Informants in client-facing roles said that it is always

more important to drink in after-hour gatherings with first-time clients than repeat clients. One of

them added that a repeat transaction meant "either that [her firm's] rate is inexpensive or they

trust us," and that is why they do not have to drink with repeat clients. This is *not* to say that the

need for conformity to visibly unpopular norms *disappears completely* when there is a more

direct indication, since there is unlikely to be a perfect signal of commitment (cf., Vaughan

1986). Nevertheless, these accounts validate the theoretical logic that conformity to visibly

unpopular norms as a signal is *most* valuable when there is little direct information about interpersonal commitment; and thus that actors conform to the visibly unpopular norm in such situations.

## *ii) When one faces an evaluation*

Drinking in after-hour gatherings also seems to be especially intense when one faces a consequential evaluation. In particular, many informants in client-facing roles said that before a contract is signed and their client has to legally bind herself to them, they need to make sure that they drink; otherwise, their client might not trust them enough to establish a legally binding relationship. For instance, a junior-level accountant said:

> Before the contract is signed...they want to make sure that we will do due diligence. So they want to go out for drinks with us to see what kind of people we are before we sign the contract. You know how they hire us because they want to defend themselves when IRS or something like that comes to town and raids them, right? They want to see [before the contract is signed] that we would do due diligence for them so that they face no risks.

Similar accounts were given for relationships within an organization as well. For instance, a mid-level manager from a pharmaceutical corporation said:

> Interviewer: Are there after-hour gatherings in which drinking is more important than others? Respondent: Well, first of all, there are no such things as after-hour gatherings that you can miss or don't have to drink...Drinking in after-hour gatherings doesn't become more or less important because of some kind of topical issue, but the importance is determined by who is coming. For example, it is especially important right before promotion decisions if my evaluators are coming.

These quotes suggest that drinking in after-hour gatherings is especially important when interactants are evaluating how trustworthy one another might be.

## *iii) Not just hierarchy but evaluation of commitment*

It is important to note that informants did not suggest that they drink *only because* their interactants are in higher positions of hierarchy. Rather, many informants alluded to promotion

evaluations as important reasons why they drink. Therefore, given how rare it is for the superiors

to be evaluated by the subordinates in Korean firms, my respondents suggested that it was much

more important to drink when circulating a glass with the superiors than subordinates. Yet, they

specifically pointed to the logic of evaluation of commitment and not the logic of hierarchy as

the driver of their intense drinking. For instance, in a conversation with a mid-level banker:

> Interviewer: When do you think people circulate their glass?
> Respondent: When they need to express loyalty?...I think people see drinking as a way by which
> they can show loyalty.
> Interviewer: When do you think loyalty shown via drinking is important?
> Respondent: For promotion decisions, a manager's opinion is really important. There are only so
> many positions for promotion, but there are many who have a good record, so it is inevitable that
> the manager's opinion is considered with importance. For this reason, people think that some kind
> of internal political relationships and personal ties influence promotion decisions.

A flipside of this conclusion is that once an employee does not want or need to signal her

commitment to her client or her superiors, she would much less need to conform to the visibly

unpopular drinking norm. That is, even if the hierarchy is equally present, insofar as the need to

signal commitment is reduced, there is less conformity. For instance, several informants

attributed lower commitment of younger workers as the primary reason why there is less

drinking in after-hour gatherings now than one or two decades ago. This inference of motives

reflects this paper's proposition that insincere conformity to the visibly unpopular norm is

specifically used as a signal of interpersonal commitment. For instance, one retiree from an

electronics firm suggested:

> One thing that changed the most is that: decades ago, one was willing to sacrifice one's time and
> everything to be able to climb up the organizational ladder, but people of younger generations
> want to have their own time rather than a little bit more of bonus...that's probably why drinking
> in after-hour gatherings now happens less frequently and less intensely.

A similar account was given by a mid-level worker in an IT firm:

> People who just joined the firm tend to separate themselves from the firm...they want to have
> their own personal time. People of older generations try to adapt their lives for their firm, but
> people who just joined, similarly to me, do not really want to, because we have our own lives. I

want to meet *my* friends. We are just more individualistic, so that's why we don't drink [in after-hour gatherings] as much.

These accounts consequently suggest that businesspeople in Korea interpret a lack of drinking in after-hour gatherings as a lack of commitment – an interpretation that in turn is likely to guide the businesspeople's strategic decisions in their own after-hour gatherings.

*Does Drinking Endow Plausible Deniability?*

Van Maanen (1991) argues that colleagues might use drinking in their after-hour gatherings to be more frank with one another. In particular, while initiating a conversation about a sensitive topic might come off as unprofessional or overly confrontational, drinking provides one with some plausible deniability of such qualities. There then should be a relationship between the permitted straightforwardness of a conversation and the amount of drinking in after-hour gatherings. Yet, many of my informants said that drinking rarely endows such plausible deniability. They said that drinking does not always lead to frank conversations; rather, after confrontational conversations take place in after-hour gatherings, political fights are often staged mornings after. Also, many informants said that while they circulate a glass with many in a given after-hour gathering, they rarely end up having real conversations with more than a few. This was especially true when one circulates a glass with someone evaluating them. Yet, despite a lack of frank conversations, circulating a glass with their superiors and drinking in front of them still took place. For instance, a mid-level banker said:

> If I'm just talking with my colleagues and not circulating my glass, someone comes and hints at me…"when will you circulate your glass with the head of the department?"…So, it is a matter of principle that I circulate my glass with someone evaluating me.

In other words, conformity to drinking does not seem to vary based on how likely one is to have a frank conversation with another. It then seems insufficient to attribute plausible deniability as the main driver of drinking in this setting.

*Consequences: Does Conformity Make One Appear as an Attractive Relationship Partner?*

The interview evidence presented thus far suggests that via drinking in after-hour gatherings, Korean businesspeople try to appear as more attractive relationship partners or organizational members. However, if they do not witness benefits from their conformity, it might not persist. Do they then actually achieve the benefits over time? While Study 2 more fully addresses this question using causal evidence, interview evidence provides consistent accounts as well. In short, informants attribute large benefits to drinking. For instance, a junior-level associate in a large corporation described how difficult it might be to get confidential yet necessary documents from other departments in his firm, but when he builds relationships through drinking, it might become easier:

> For instance, department X has some document [that we need], and the official procedure is to file an administrative request...but when someone knows someone from drinking together, he can do some prodding to get that document.

Another informant who used to work in a major electronics company as a middle manager said:

> Relationships from drinking together in after-hour gatherings are actually good for me. In my case, in my team [at previous company], I used to be in charge of ordering the team's office supplies, and some are not exactly ones that I can ask for without any hesitation. For instance, there is no clear standard for the computer keyboard or mouse that I'm going to use personally, so those who make purchases have to make a judgment call. So when they are deciding whether to buy or not, it probably is better if you are someone with the relationship.

Moreover, as much as drinking in after-hour gatherings might be rewarded, several informants reported that nonconformity elicits negative consequences to individuals and organizations. All informants said that while drinking in after-hour gatherings is pervasive, not

everyone drinks. And most of them listed numerous accounts documenting penalties for not

drinking. For instance, in a conversation with a mid-level banker:

> Interviewer: Is there anyone who does not drink in after-hour gatherings?
> Respondent:...They don't force a non-drinker to drink. There are a few nondrinkers – one or two
> [out of 15]? The head of the department and the head of the team treat them as outsiders. Not just
> in after-hour gatherings, but also during work or when they need to assign work for something,
> the non-drinkers are never the first ones to be called on. When I wasn't drinking back in
> September [because of health reasons], they kept saying "when are you going to drink?" I felt like
> they were saying "if you don't drink, you will be like them."

It is worth noting that the non-drinkers were described as "outsiders" by this informant. Other

informants similarly used words such as "not loyal" or "untrustworthy" to describe how non-

drinkers were perceived in their organizations. Such benefits of relationships built through

drinking together seem to manifest themselves in interorganizational relationships as well. A

junior-level consultant in a credit agency said that he is more likely to follow up with a client

with whom he built relationships via drinking, even after their contractual obligations have been

fulfilled:

> When we have personal relationships that were built via drinking together...we can call one
> another after the project is done, and say "how are you doing" or "are there any problems." But I
> wouldn't have this kind of conversation with those with whom I just had business relationships.

Interestingly, informants in client-facing roles reported fewer incidences of

nonconformity in their after-hour gatherings. But such variation seems to be precisely because

workers in client-facing roles try to make sure that they appear committed in their

interorganizational relationships filled with more uncertainties. For instance, a mid-level

manager of a *chaebol* who worked with consultants said that her department "brought only the

best drinkers to after-hour gatherings in order to make sure that everyone [from her department]

could drink intensely." Cognizant of the signaling value, consequently, many informants said

that even though they want to avoid drinking, they still prefer drinking with their clients or

consultants to build relationships. For instance, a junior-level accountant said:

> My clients can make me do analysis on so-and-so, an analysis that is complex and takes a long time. But after doing the analysis, I realize that they made me do it just to make us go through hell. They see the analysis, and they say "ah, good work. But we don't need it anymore." And then they move on. It took us one whole week to do that analysis! They can easily do that with the intent of testing us. Then, if we want to not go through hell and if we want to finish the project on time, the relationship with them becomes super important. By drinking in front of them as they expect to see…this is all good for us in the end as well.

Similarly, a senior-level employee in a construction company said:

> When I first went to the field, the manual workers tested me. They were thinking, you have lots of education under your belt, but what can you really do? One test was going out to see how much I could drink. Another was going up to the construction crane and walking back and forth…so the best solution is to drink when they suggest that I drink.

Ultimately, these accounts show that some sacrifice is often expected as a demonstration of one's interpersonal commitment; and that many Korean businesspeople recognize drinking in after-hour gatherings as the most convenient solution, potentially further perpetuating its persistence.

### *Study 1 Discussion*

Consequently, these accounts suggest that businesspeople in Korea have strategic motives for drinking despite their individual and collective distastes, insofar as it is perceived as the normative standard. But this is not to say that everyone had the same opportunities to engage in this strategic action. In particular, women often suffered from not having such opportunities. They often expressed that they have difficulties showing their loyalty to their interactants and group members because they are often not included in the ritual of glass-circulating; and this difference in opportunities seems to result in different outcomes. For instance, a senior-level female employee in a construction company said:

> There are many risks to which women get exposed in after-hour gatherings. Sexual assault, for instance, either by those within our team or by clients…Men don't have that problem…So when women are at an after-hour gathering, we think "ah, should I not be here?" And male colleagues might nudge us to leave early, saying something like "it's uncomfortable that you are here, so go home at an appropriate time."…so we get isolated that way.

Lastly, it is worth noting that even when many informants described unpopularity around the drinking norm, they called it a "necessary evil" in their workplace. That is, even though they appreciated the functional benefit of violating one's preference, most informants recognized it as a suboptimal coordination standard and wished that they could opt out to an alternative way of signaling their commitment. Several of them told me at the end of the interviews that they really hope that I – as a researcher – can come up with ways in which to abolish drinking in after-hour gatherings. Their attitude toward drinking in after-hour gatherings then might best be summarized by the following quote by a junior-level engineer in a small electronics company:

> I think [drinking in after-hour gatherings] is a very bad custom. People tend to think that drinking excessively will show one's loyalty and one's sincerity, but everyone will die if everyone keeps doing this! It's not good for our health either. Every person is human capital for the company…It really is a bad custom.

Consequently, these accounts then highlight that the Korean businesspeople coordinate on a suboptimal normative standard. That is, conformity to the visibly unpopular drinking norm is driven by the reluctant need to coordinate on a point from which everyone prefers to deviate.

## Study 2: Vignette Experiments

But does insincere conformity here actually work as a signal of interpersonal commitment? While evidence from Study 1 showed that Korean businesspeople strategically conform to the visibly unpopular drinking norm in after-hour gatherings, a complementary study is needed to more directly test what it is about conformity to visibly unpopular norms that signals commitment to one's interactants. On one hand, it might be *insincere* conformity to the visibly unpopular norm. On the other hand, it might be *sincere* conformity, as for the visibly unpopular norm as well as popular norms. My theory predicted the former, since the inference of motive is then isolated to be that of acting on behalf of the collective. Even though interview evidence

from Study 1 was consistent with this account, a more direct test on the effect of insincere

conformity (and the effect of sincere conformity) is needed.

Also, a causal test is needed to establish this paper's theory as a sufficient mechanism

driving conformity to visibly unpopular norms. While informants in Study 1 allude to the

signaling value as the driver of their insincere conformity, they might be misattributing. Also,

while informants in Study 1 allude to the perception of *interpersonal commitment* as the outcome

of their insincere conformity, they might be misattributing on this as well. Therefore, I need to

causally isolate the signaling effect of insincere conformity on the perception of commitment.

Consequently, the two-fold goal of Study 2 is (i) testing whether it is specifically insincere

conformity that increases the perception of one's commitment; and (ii) testing whether the

insincere conformist's higher likelihood of being preferred is driven by a perception of

commitment and not just any positive valence. This two-fold goal is achieved via experimentally

testing specific empirical implications of the theory developed above, as summarized by the

following:

*Hypothesis 1: In a setting where audiences are assessing potential relationship partners,*

*preference for the insincere conformist to the visibly unpopular norm is greater than that*

*for the nonconformist.*

*Hypothesis 2: The higher preference for the insincere conformist is explained by*

*audience perception of the conformist's interpersonal commitment.*

Note that a test for Hypothesis 2 can validate conditions under which conformity to visibly unpopular norms is less useful (therefore when it might stop). This paper's theory implied that drinking in after-hour gatherings is useful as a *signal* of interpersonal commitment; therefore, by definition, when there is more direct information about one's commitment, drinking should not enhance one's perception of interpersonal commitment as much. Therefore, while some informants in Study 1 suggested that drinking is less needed when there are more "data" about the person, I design a more directly causal test in Study 2. Providing a way in which conformity to visibly unpopular norms is less needed is important for practical reasons as well, insofar as persistence of unpopular norms might decrease worker satisfaction and increase turnover (Cordes and Dougherty 1993).

Finally, it is worth clarifying that since Study 2 is designed to complement Study 1, Study 2 focuses on testing what Study 1 could not test. That is, Study 2 is *not* designed to tap into individuals expressed motives for conformity, since Study 1 provided more direct evidence for that.

### Study 2 Data and Methods

I provide empirical validation of these hypotheses with online survey vignette experiments. 500 Korea-based subjects[11] who have previously worked in Korea were recruited via Korean *SurveyMonkey*[12] in November of 2016. The experiment was conducted in Korean.

---

[11] While 500 subjects were paid to participate, an error occurred on the platform while collecting one subject's response and it was henceforth excluded from the analysis.

[12] This is a similar platform to Amazon Mechanical Turk in the U.S. To my and platform administrators' knowledge, no studies exist that test whether subjects on this platform behave the same as those in the lab, like those for Amazon Mechanical Turk in the US (e.g., Berinsky, Huber, and Lenz 2011; Buhrmester, Kwang, and Gosling 2011). The platform is most often used by retail and advertising companies to do market research on consumers' opinions.

Each subject was paid 3,100 wons (equivalent to about $2.60) for participating. Since drinking in

after-hour gatherings is a norm specific to the Korean context, sample recruitment was focused

on Koreans who have had work experience in Korea. The online experimental setting enabled

me to sample subjects who can make contextualized judgments about familiar social interactions,

making it an "online field experiment" (Parigi, Santana, and Cook 2017) where subjects make

decisions that they would have faced outside the vignette setting. However, subjects in the online

platform may pay less attention than those participating in laboratory settings. Therefore, as has

become conventional, I addressed this challenge in two ways. First, I broke down the experiment

into "steps," so that subjects receive a small amount of information at each click. Second, I asked

them a series of attention questions. By doing so, I could track whether they were paying

attention and understanding my instructions as intended (Mason and Suri 2011). Lastly, about a

half of the sample was female; and about 48% of the sample was in their 30s.[13]


## *Study 2 Conditions*

Upon entry, subjects are randomly assigned to one of three sets of conditions; and in each

set of conditions, subjects choose who would be a more preferable business partner between two

candidates. In all conditions, subjects see a conformist with unknown preference toward drinking

(i.e., *Baseline conformist*) as a candidate. Since the drinking norm is visibly unpopular as shown

in Study 1, a *Baseline conformist* is assumed to dislike drinking and conformity implicitly

violates his preference. This category is likely the most common case of conformity, since a

conformist's preference to the visibly unpopular drinking norm is most often not explicitly

---

[13] There was no interaction effect between any demographic variables collected and the main treatment
(possibly for how the questions were asked, which is described below), so I do not discuss it further.

known by interactants (even though it is implicitly assumed to be low). I also confirm that

subjects see a *Baseline conformist* as an insincere conformist through manipulation checks.

In addition this *Baseline conformist*, subjects in each set of conditions see one other actor.

In one set of conditions (*Suffering conformist vs. Baseline conformist* conditions), subjects

additionally see a *Suffering conformist* who is explicitly known to dislike drinking (i.e., one

whose conformity is explicitly insincere). In another set of conditions (*Enthusiastic conformist

vs. Baseline conformist* conditions), subjects see an *Enthusiastic conformist* who is known to like

drinking (i.e., one whose conformity is explicitly sincere). In the last set of conditions

(*Nonconformist vs. Baseline conformist* conditions), subjects see a *Nonconformist* who is known

to dislike drinking and does not drink.[14] Subjects in each set of conditions then express how

much they prefer a *Baseline conformist* versus the other actor (i.e., *Suffering conformist*,

*Enthusiastic conformist*, or *Nonconformist*). Table 3 presents the overall layout of these

conditions.

Table 3 around here

Lastly, Hypothesis 2 suggests that the higher preference for an insincere conformist

should be explained by a perception of interpersonal commitment. Therefore, in addition to

varying the degree of conformity and the degree of sincerity, experimental conditions of Study 2

vary on the amount of direct information an audience receives on one's commitment. If the

greater preference for an insincere conformist is driven by perceptions of commitment

specifically, the preference should become much less salient after direct indications of

commitment are given. Therefore, in one set of conditions, subjects are told no direct

---

[14] A *Nonconformist* is designed to dislike drinking as well as refuse drinking, so that between-subject tests (described more below) can be more parallel. Additionally, there is no condition where one likes drinking but does not conform, since such a scenario is neither plausible nor theoretically interesting.

information about potential partners' commitment (*No information* conditions). In another set of

conditions, subjects are told direct information about their commitment, one's commitment

apparently higher than the other's (*Direct information* conditions). These conditions then also

allow me to test whether insincere conformity has less strategic value when there is more direct

information about one's commitment. Information about the two candidates for a relationship

partner is counterbalanced in each of *Direct information* conditions, which is why those

conditions have about twice as many subjects as do *No information* conditions. These conditions

are again summarized in Table 3.

## *Study 2 Vignette*

I will first lay out the *Nonconformist vs. Baseline conformist* condition in the *No*

*information* condition. Afterward, I will elaborate on how other conditions differ. The flow of

the vignette for different conditions is summarized in Figure 3. Note that subjects in *No*

*information* conditions were taken straight to step (b) in Figure 3, without seeing step (a).

Figure 3 around here

### *Nonconformist vs. Baseline conformist in the No information condition*

After consenting to participate in the experiment, subjects were told that they would read

a story about a large corporation's strategy department from the perspective of Associate

Daewoo Kim.[15] Subjects were further told that they would later be asked what decision they

think Associate Kim would have made at the end of the story (more below on the questions

subjects were asked).

---

[15] All names used in the experiment are generic male Korean names.

After the introduction, subjects were told the following story in two consecutive screens. They were told that in March 2015, a large corporation in Korea reorganized its employees across different departments; and that Associate Kim was assigned to the strategic management department. Subjects were further told that as part of the reorganization, employees would increasingly work with those outside of their teams. Subjects were then told that Associate Kim's team was given a project on which it has to collaborate with another team; and that Associate Kim has been given the responsibility of choosing with which team to collaborate. Subjects were further told that choosing with which team to collaborate is an important decision; and that successful collaboration usually requires team members to strive for collective success over their individual performance. Subjects were then asked some attention questions.

Afterward, subjects were told the following story in two consecutive screens. They were told that Associate Kim decided to choose one of two teams that were then not working on any other project – Associate Hanmin Park's team or Associate Sangchul Lee's team. Subjects were further told that employees at an associate level were crucial in facilitating collaboration between two teams; but that Associate Kim did not know either Associate Park or Associate Lee personally or professionally; and therefore, Associate Kim asked his colleague at HR department if he could take a look at their performance evaluations. Subjects were however told that the colleague in the HR department said that no one outside of the HR could look at employees' evaluations. Subjects were then asked attention questions.

Afterward, subjects were told the following story in four consecutive screens. Subjects were told that (i) an after-hour gathering was scheduled to finish off the reorganization; and (ii) at the gathering, while the food was being ordered, Associate Kim saw another associate suggest circulating a glass with Associates Park and Lee. Afterward, subjects were told that Associate

Kim remembered hearing something about Associate Park a few days ago. Subjects were told that Associate Kim had heard from his colleague that "Associate Park really dislikes drinking in after-hour gatherings."[16] However, when asked, his colleague said that he did not know whether or not Associate Lee likes drinking. In other words, subjects were told that Associate Park likes drinking, but they were not told anything about Associate Lee's preference for drinking. Subjects were then asked attention questions and manipulation checks on what they infer each person's preference for drinking to be.

Lastly, subjects were told whether or not Associates Park and Lee drank. Subjects were told that Associate Kim saw Associate Lee bottom up the glass, but that Associate Park refused to take a shot. In other words, Associate Park is a *Nonconformist*, and Associate Lee is a *Baseline conformist*. When Associate Park was suggested again to drink, subjects were told that Associate Park said that he would take the glass but would not drink.[17] Afterward, subjects were told that after a while, members of the department finished eating and left after saying goodbyes to one another.

*Suffering conformist vs. Baseline conformist in the No information condition*

Subjects in *Suffering conformist vs. Baseline conformist* conditions in the *No information* conditions were told the exact same story as in *Nonconformist vs. Baseline conformist* condition in the *No information* condition, except for one detail – subjects were told that Associate Park (*Suffering conformist* in this condition) drank.

---

[16] Before manipulating preferences this way, I asked a few informants from Study 1 how common it is for them to know someone else's preference for drinking; and they confirmed that neither knowing someone's preference nor not knowing it comes across as strange or signifies other characteristics about him or her. An analogous example in the US context might be knowledge about one's dietary restrictions (e.g., vegetarianism).

[17] As shown in an informant's quote in Study 1, this is a polite way of refusing a drink.

*Enthusiastic conformist vs. Baseline conformist in the No information condition*

Subjects in *Suffering conformist vs. Baseline conformist* conditions in the *No information* conditions were told the exact same story as in as in *Nonconformist vs. Baseline conformist* condition in the *No information* condition, except for two details. First, subjects were told that when Associate Kim was talking with his colleague about how much Associate Park likes (or dislikes) drinking, Associate Kim heard that "drinking is something that Associate Park looks forward to in those after-hour gatherings." Associate Kim however did not hear whether or not Associate Lee likes drinking, as in all other conditions. Second, subjects were told that Associate Park drank.


*Each condition in the Direct information conditions*

These three conditions have parallel conditions in the *Direct information* conditions. The only difference between the *No information* conditions and the *Direct information* conditions is that when Associate Kim asked his colleague in the HR department to see performance evaluations of Associates Park and Lee, Associate Kim was able to see their evaluations. Subjects were told that Associate Park (i.e., *Suffering conformist, Nonconformist,* or *Enthusiastic conformist*) was recorded as "highly capable" and "committed to team he is assigned to" – among other descriptors – and given 4 out of 5 for a competence score and 5 out of 5 for a commitment score, given by about 60 colleagues. Subjects were told that Associate Lee (i.e., *Baseline conformist*) was recorded as "highly capable" and "sometimes hard to work with" – again, among other descriptors – and given 4 out of 5 for a competence score and 2 out of 5 for a commitment score, again given by about 60 colleagues. In particular, the descriptors as well as scores for the two Associates' competence were kept the same, while descriptors and scores for

their commitment differed.[18] These conditions were counterbalanced, so that the other half of

subjects in *Direct information* conditions saw Associate Lee as more committed. Except for this

part, subjects in *Direct information* conditions were told the exact same story of their respective

conditions in *No information* conditions.

### Study 2 Dependent Variables

After they finished reading the story, subjects in all conditions were told that "A week

later, Associate Kim decided with which team to collaborate," and asked the following question

"With whose team do you think Associate Kim chose to collaborate?" It is worth noting two

things about this dependent variable.

First, subjects were asked for their third-order beliefs – as were informants of Study 1 –

because this is a measure of social consensus beliefs that are more likely to be acted upon than

privately held beliefs (Ridgeway and Correll 2006:433-434; Correll et al. 2017). Also, while the

primary objective of Study 2 is to see how effective insincere conformity to the visibly

unpopular norm is as a signal, insofar as the proposition pertains to why conformity to a visibly

unpopular norm continues, beliefs about others' beliefs are critical to this paper's theory.

Therefore, subjects were asked to evaluate the statements "Associate Kim chose to collaborate

with Associate Park's team" and "Associate Kim chose to collaborate with Associate Lee's

team" on scales ranging from 1 (Very unlikely) to 10 (Very likely).

---

[18] Recall that in Study 1, where many informants reported that they would drink more in front of their evaluators before promotion decisions. In order to prevent subjects from thinking that Associates Park's or Lee's performance evaluations resulted from impressions formed in after-hour gatherings, descriptors that attributed their competence and commitment were added so that the performance evaluation appears more "objective" and based on work performance. These evaluations were designed based on real performance evaluations used in Korean firms.

Second, because I was primarily interested in whether an insincere conformist is preferred *in comparison to* a nonconformist, the question asked which team Associate Kim would have chosen between the two. This question therefore frames the perceived preference for one candidate to be *dependent* on the perceived preference for the other candidate; thus, the primary variable of interest is the *difference* between the two likelihoods (e.g., differential perceived preference for Associate Park's team over Associate Lee's team): since the perceived preference for one candidate is *dependent* on the perceived preference for the other candidate, the absolute level of perceived preference for Associate Lee is likely to change in different conditions, even though he is a *Baseline conformist* in all conditions. Thus, the difference between the two is used as the relative preference for Associate Park (i.e., *Suffering conformist, Nonconformist,* or *Enthusiastic conformist*) over Associate Lee (i.e., *Baseline conformist*). Lastly, after responding to these statements, subjects were asked for their third-order beliefs about competence ("*How competent at task do you think Associate Kim would think of Associate Park?*") and commitment ("*How committed to other employees do you think Associate Kim would think of Associate Park?*") of each candidate. These questions are to see whether the higher perceived preference for the candidates was driven specifically by a perception of commitment, while a perception of capability (as well as a perception of commitment) might also enhance one's preference as a relationship partner (Gould 2002; Correll, Benard, and Paik 2007). After subjects answered additional questions, including free response items, they were thanked and given a code to be paid.

*Study 2 Results and Analyses*

*Confirming that a* Baseline Conformist *Is Perceived to be an Insincere Conformist*

Before showing evidence for hypothesis-testing, I first establish that subjects indeed infer

a *Baseline candidate* – whose preference for drinking is not explicitly known – to dislike

drinking. Confirming that this assumption holds true is especially important, since the theory

suggests that the conformity to the visibly unpopular is assumed to be insincere. And results

from manipulation checks validate this assumption. When asked "How much do you think [the

candidate] enjoys drinking?" (1=Not at all to 10=Very much), subjects perceived a *Baseline

conformist* to enjoy drinking (M=1.60) significantly less than an *Enthusiastic conformist*

(M=6.91; t=33.71; p<0.001), but not significantly more or less than a *Suffering conformist*

(M=1.39; t=1.09; p=0.28). These results then show that both *Baseline conformist* and *Suffering

conformist* are seen as insincere conformists – the former implicit and the latter explicit –

supporting the theoretical logic and making the below analyses appropriate.


*Validating Hypothesis 1*

Hypothesis 1 predicted that an insincere conformist to the visibly unpopular drinking

norm would appear more preferred as a relationship partner than a nonconformist. Since this

prediction is about when there is little information about one's commitment, tests for Hypothesis

1 are done within *No information* conditions. And results validate this prediction. One way of

testing this hypothesis is by doing a within-subject comparison in the *Nonconformist vs. Baseline

conformist* condition, since I then can test whether an implicitly insincere conformist (i.e.,

*Baseline candidate*) is perceived to be preferred over a nonconformist. This within-subject

comparison indeed confirms that a *Baseline conformist* is perceived to be preferred (M=6.50)

over a *Nonconformist* (M=5.21; t=3.52; p-value<0.001). Figure 4 shows this finding.

Figure 4 around here

Another way of testing Hypothesis 1 is by running between-subject comparisons: by

confirming whether relative preference for a *Suffering conformist* (compared to a *Baseline*

*conformist*) is greater than that for a *Nonconformist* (compared to a *Baseline conformist*). This

comparison then tests whether an explicitly insincere conformist is perceived to be preferred

over a nonconformist. This between-subject comparison again confirms that the perceived

relative preference for a *Suffering conformist* is significantly greater (M=-0.06) than that for a

*Nonconformist* (M=-1.29; t=2.24; p-value<0.05). By contrast, there is no evidence that the

perceived relative preference for an *Enthusiastic conformist* (M=-0.95) is larger than that for a

*Nonconformist* (M=-1.29; t=0.53; p-value=0.60). These results then provide evidence that

insincere conformity in particular (and not just any conformity) increases the likelihood of being

preferred as a relationship partner.

Other results further confirm that it is specifically the insincerity of conformity that

increases one's perceived preference as a relationship partner and not just any conformity. The

within-subject comparison in the *Enthusiastic conformist vs. Baseline conformist* condition

shows that a *Baseline conformist* appears more preferable (M=6.49) compared to an *Enthusiastic*

*conformist* (M=5.53; t=2.45; p-value<0.05). This comparison most directly compares

conformists of two varying degrees of sincerity; and it shows that the insincere conformist is

even preferred over the sincere conformist. This effect was not specifically hypothesized, since

the inference of motive for the sincere conformist is *ambiguous* and not that of outright lacking

commitment; but it is certainly consistent with the overall theory. Lastly, and unsurprisingly, the

same within-subject comparison in the *Suffering conformist vs. Baseline conformist* condition shows no evidence that either is significantly preferred over another (M for *Suffering conformist*=5.83; M for *Baseline conformist*=5.89; t=0.16; p-value=0.87), since both of them are insincere conformists, the former explicit and the latter implicit. These analyses then further show that insincere conformity to the visibly unpopular drinking norm (and not any conformity to the visibly unpopular drinking norm) leads to the greater preference as a relationship partner. Again, Figure 4 summarizes these findings.

*Validating Hypothesis 2*

Hypothesis 2 predicted that the greater preference for an insincere conformist to the visibly unpopular norm would be explained specifically by a perception of commitment. There are two ways to validate this prediction: (i) by validating that the effect of insincere conformity to the visibly unpopular drinking norm in *No information* conditions is primarily explained by a perception of commitment (but not by a perception of competence); and (ii) by providing a more direct indication of commitment (i.e., using *Direct information* conditions) and validating that the difference between insincere conformity to the visibly unpopular drinking norm and nonconformity then goes away. And results from both analyses validate Hypothesis 2. Model 1 in Table 4 presents the OLS regression of the positive and significant effect of being a *Suffering conformist* on preference as a relationship partner. Model 2 shows that being a *Suffering conformist* also increases the perception of commitment; moreover, when Model 4 includes the perception of commitment to the same model as Model 2, the effect of the perception of commitment on preference as a relationship partner is positive and significant, and there no longer is evidence that being a *Suffering conformist* directly increases preference as a

relationship partner. These models then directly validate Hypothesis 2, i.e., that the higher

preference for the insincere conformist is primarily explained by the perception of commitment.

By contrast, Model 3 shows no evidence that being a *Suffering conformist* increases the

perception of competence; and the positive and significant effect of the perception of

commitment from Model 4 holds even when Model 5 controls for the perception of competence.

These models then confirm that the higher perceived preference for an insincere conformist to

the visibly unpopular drinking norm is mainly driven by the enhanced perception of commitment

and not just any positive valence (e.g., perception of competence).[19]

Table 4 around here

Another way of testing whether the main effect is driven by perceived commitment is by

seeing whether the relative preference for a *Suffering conformist* (over a *Baseline conformist*) is

still greater than the relative preference for a *Nonconformist* (over a *Baseline conformist*), after

both the *Suffering conformist*'s and the *Nonconformist*'s commitment has been shown to be

greater than that of a *Baseline conformist* via more direct information (i.e., in *Direct information*

conditions). This difference should disappear, if the effect of insincere conformity to the visibly

unpopular drinking norm was driven by perceived commitment. Results confirm this predicted

pattern, as shown in Figure 5. The comparisons in *Direct information* conditions indicate that the

relative preference for a *Suffering conformist* is no longer significantly greater (M=1.92) than the

relative preference for a *Nonconformist* (M=1.74; t=0.30; p-value=0.77), by contrast to the

---

[19] Despite wide use in the field, I do not claim that this main relationship is "mediated" by the perception
of commitment because my experimental design only manipulates how sincere conformity is and does not
directly manipulate how committed the potential relationship partners are (see Gelman and Hill 2006:188-
194). Thus, I use the language "explain" to indicate that the perception of commitment explains more
variance and thus is more responsible (than the perception of competence) for increasing the preference as
a relationship partner within the context.

significant effect for the same comparison in *No information* conditions.[20] Also, within-subject

comparisons show that subjects are significantly more likely to prefer a *Nonconformist*, *Suffering*

*conformist*, or *Enthusiastic conformist* over a *Baseline conformist*, when the former categories of

actors are directly indicated to be more committed to their team, while the opposite was true in

*No information* conditions. Counterbalanced conditions (i.e., where subjects are given direct

information that commitment of a *Baseline conformist* is higher) show substantively the same,

but reversed pattern. In other words, in those counterbalanced conditions, a *Baseline conformist*

is preferred over a *Nonconformist*, *Suffering conformist*, or *Enthusiastic conformist* and there no

longer is evidence for differences in relative preference among the three conditions.

<div align="center">Figure 5 around here</div>

### *Study 2 Discussion*

Study 1 established that Korean businesspeople have strategic motives for conforming to

the visibly unpopular norm around drinking in after-hour gatherings. Study 2 directly and

---

[20] The non-significant effect of insincere conformity to the visibly unpopular drinking norm within *Direct information* conditions however does not necessarily mean that the signaling effect of insincere conformity to the visibly unpopular drinking norm *disappears*: it only means that there no longer is evidence for the signaling value of insincere conformity. In order to confirm that the effect of insincere conformity is negligible or effectively "null" when there is direct information, I ran additional tests where I compare the signaling effect of insincere conformity in *Direct Information* conditions to the "true" effect, i.e., the signaling effect of insincere conformity when there is no information about commitment (Rainey 2014). In order to do so, I estimated the 95% confidence interval of the difference between the relative preference in the *Suffering conformist vs. Baseline conformist* condition and that in the *Nonconformist vs. Baseline conformist* condition, within *Direct information* conditions. After that, I checked whether that confidence interval overlaps with the "true" effect of insincere conformity to the visibly unpopular drinking norm (relative to nonconformity), as shown in the between-subject in the same pairwise comparison within *No information* conditions. The results show that the 95% confidence interval ranges from -1.39 to 1.03 and does not include the true effect (1.23), confirming that insincere conformity to the visibly unpopular drinking norm has a negligible signaling effect when commitment has been more directly indicated. The negative value in the confidence interval indicates that the relative preference for a *Nonconformist* is greater than the relative preference for a *Suffering conformist*; the positive value indicates the opposite.

causally tested whether such strategic actions are effective. Finding out whether the conformity's insincerity (and not conformity in and of itself) drives the perception of commitment is important for this paper's theory, since it predicted that audiences would prefer specifically *insincere* conformists over nonconformists. The results confirmed that an insincere conformist to the visibly unpopular drinking norm is preferred as a relationship partner over a nonconformist but there is no evidence that a sincere conformist enjoys the same, showing that it is specifically insincere conformity that increases the level of preference as a relationship partner.

Another noteworthy finding from Study 2 is about conditions under which the commitment-signaling effect might be less salient. The second test for Hypothesis 2 showed that once one's commitment has been demonstrated more directly via performance evaluations, insincere conformity to the visibly unpopular drinking norm no longer led to the greater perceived preference as a relationship partner. These findings then confirm that the signaling value of insincere conformity to visibly unpopular norms is less salient once there are more direct indications of the conformist's interpersonal commitment. The experimental manipulation in the vignette thus provides one practical and plausible way in which conformity to the drinking norm might become less salient (hence conditions under which conformity to a visibly unpopular norm might stop). In particular, by promoting evaluations for which one's level of commitment is strictly based on more "objective" criteria, organizations might be able to minimize persistence of norms from which they and individuals prefer deviating.

## GENERAL DISCUSSION

Existing approaches see persistence of an unpopular norm as a misperception problem: individuals misconstrue a norm as representing the group members' values and preferences; and

conform so as to feign endorsement. The important implication of these approaches is that removing the misperception is sufficient to stop conformity to unpopular norms. But this paper's theory and evidence suggest to the contrary. That is, the norm's visible unpopularity provides a strategic opportunity to signal one's willingness to violate one's preference and act on behalf of others. Complementary evidence from interview and experimental studies supports for this theory, using the visibly unpopular norm prescribing excessive drinking in after-hour gatherings in Korea. Then, insofar as actors recognize this signaling value, individuals might coordinate on the suboptimal normative coordination point even though they prefer to escape it.

I now extend this paper's empirical results to the broader context by elaborating on how costly signals here operate and how individuals might try to feign their distaste for the norm so as to appear committed (and limits of such performative actions). I then conclude by clarifying implications of this paper's theory and findings.

## Costly Acts for Commitment to Interactants (and Not Norms Themselves) and "Performing" Insincerity of Conformity (and Limits of Such Performance)

I have argued that the effectiveness of insincere conformity to a visibly unpopular norm hinges on the apparent costliness. This idea builds directly on existing theories of costly signals (e.g., Spence 1973, 1981; Akerlof 1976; Landers et al. 1996; Henrich 2009; Willer 2009), which suggest that costliness of an act enhances sincerity of the claim. Then, together with existing theories, my theory provides a useful way in which costly acts enhance one's commitment to interactants (as well as one's commitment to the norm). Consider Willer (2009), who suggests that one's costly participation in collective action enhances the perception that one is committed to the cause of the collective action (Centola, Willer, and Macy 2005; Willer et al. 2009; cf., Kim

and Zuckerman Sivan 2017). Similarly, Power (2017:83) suggests that costly acts are used to "[convey] information about the…signaler's commitment to the…precepts of the community." By contrast, my theory suggests that the costly conformity is used not to signal commitment to the "precepts" or a widely endorsed value of the group, but used in violation of the widely shared preference to signal *interpersonal* commitment. Thus, while insincerity of conformity to popular norms would raise eyebrows as to one's commitment to the group's shared values, insincerity of conformity to visibly unpopular norms *enhances* one's perception of interpersonal commitment.

Then, insofar as one's conformity to visibly unpopular norms is motivated by the desire to achieve the signaling value, perhaps the most preferable strategy is to *feign discontent for the visibly unpopular norm without actually violating one's preference*. In fact, individuals do seem to engage in performative local actions to enhance their perception of commitment. Recall the accounts that some informants provided that explicitly specified why their interactants might take their drinking as a signal of commitment: e.g., "It's like, as they see me bottom-up the drink and cringe, they feel 'ah, this person is really willing to obey me.' Like, doing a bottom-up even though alcohol is bitter." Based on this recognition, this informant then might *pretend* as though he does not like drinking. Similarly, another account suggested that Korean businesspeople might try to appear as if they do not enjoy drinking, even though they actually do not mind it: e.g., "What I did regret in my two years [in my last workplace] is that…I had a reputation for [enjoying] drinking." These accounts then imply that Korean businesspeople might try to overplay their sacrifice in order to appear committed to their interactants. Another implication is that the perception of the norm's unpopularity might be reinforced through performative local actions, since it is in individuals' interests to actively show that they disapprove of the norm, even if they do not (cf., Centola et al. 2005; Willer et al. 2009).

Other empirical accounts that raised the original question of this paper also document such local actions through which individuals try to overplay their sacrifice. For instance, Weeks (2004) argued that the corporation's culture encompassed not only conforming to visibly unpopular norms embedded in its culture but also complaining about the norms and culture of the corporation. It is also easy to observe such self-deprecations on one's own norms and culture in other anecdotal examples, even though members of the community actually seem to pride themselves on conforming to such norms (e.g., phrases such as "Where Fun Goes to Die" at University of Chicago; "IHTFP" at Massachusetts Institute of Technology).

One implication of these performative actions is that the apparent costliness can no longer be trusted and thus insincere conformity cannot be a credible signal of interpersonal commitment. But the possibilities for such performative actions are not tantamount to thinking that one can *always* make one's normative conformity appear costly. First of all, there might be some unchangeable facts that make conformity appear more costly. For instance, many of my informants suggest that those with spouses and kids are often deemed to violate private interests when drinking in after-hour gatherings than those without, since the latter is presumed to have more disposable time. Therefore, performative local actions might go only so far in enhancing the perception of personal sacrifice.

Moreover, asserting one's distaste for a norm might very well be perceived as "cheap talk" (cf., Crawford and Sobel 1982; Kreps and Wilson 1982; Milgrom and Roberts 1982). Possibilities for such cheap talk might be higher when interactants perceive there to be good reason to feign one's commitment (e.g., when stakes for betrayal are high). Therefore, one's distaste for conforming has to be *unambiguous* enough, and such requirement of unambiguity might make norms that are glaringly unpopular more likely to persist when the need for

commitment is high. Indeed, many visibly unpopular norms that are used to signal commitment in the most consequential contexts are often those that are *undoubtedly* disliked. Examples of such glaringly unpopular norms come from groups like gangs and fraternities (Gambetta 2011), where individuals often incur physical harms (e.g., cutting off one's finger) to demonstrate their commitment to their groups. These examples of a visibly unpopular norm then show that the effectiveness of performative actions might be narrow, since suspicions for ulterior motives might be equally high when performative actions are possible.

## IMPLICATIONS & CONCLUSION

Extant literature propose that exposing a norm's unpopularity is sufficient to stop conformity (Miller and Prentice 2016). By contrast, I argued that doing so might not be sufficient, since a norm's visible unpopularity provides a strategic opportunity to demonstrate one's commitment to the interactants. Ironically, conformity to visibly unpopular norms might be able to signal interpersonal commitment precisely because one's conformity is assumed to be insincere and based on sacrifice; otherwise, the conformity might appear to originate from one's endorsement and thus fail to signal commitment to the interactants.

This conclusion then provides both discouraging and hopeful takeaways for those in pursuit of abolishing norms from which groups and individuals prefer deviating. The strategy of informing individuals of the norm's unpopularity might further *encourage* (insincere) conformity as a symbolic gesture of interpersonal commitment. Insofar as groups try to abolish those norms, then, they should thus be cautious in taking such steps (cf., Munsch et al. 2014). Results from Study 2 suggest that the more effective solution might be to provide direct information on individuals' commitment via more "objective" evaluation. But I am cautious not to suggest such

a strategy as a silver bullet, as the "objective" performance evaluation on one's commitment might also be influenced by one's conformity to visibly unpopular norms at a previous period. An implication then is that organizations should be mindful of the (individually useful) signaling value of conformity to visibly unpopular norms if they try to prevent persistent conformity to the collectively undesirable norm.

# REFERENCES

Akerlof, George. 1976. "The Economics of Caste and of the Rat Race and Other Woeful Tales." *The Quarterly Journal of Economics* 90 (4): 599–617.

Aronson, Elliot, and Judson Mills. 1959. "The Effect of Severity of Initiation on Liking for a Group." *The Journal of Abnormal and Social Psychology* 59 (2): 177–81.

Asch, Solomon E. 1951. Effects of Group Pressure on the Modification and Distortion of Judgments. In *Groups, Leadership and Men*, edited by H. Guetzkow, 177–190. Pittsburgh, PA:Carnegie Press.

Arrow, Kenneth. 1970. "Political and Economic Evaluation of Social Effects and Externalities." *NBER*, January, 1–30.

Axelrod, Robert. 1986. "An Evolutionary Approach to Norms." *American Political Science Review* 80 (4): 1095–1111.

Azoulay, Peirre, Nelson P. Repenning, and Ezra W. Zuckerman. 2010. "Nasty, Brutish, and Short: Embeddedness Failure in the Pharmaceutical Industry." *Administrative Science Quarterly*, 55(3): 472-507.

Bailyn, Lotte. 2006. *Breaking the Mold: Redesigning Work for Productive and Satisfying Lives.* Ithaca, NY: Cornell University Press.

Bendor, Jonathan, and Piotr Swistak. 2001. "The Evolution of Norms." *American Journal of Sociology* 106 (6): 1493–1545.

Bernheim, B. Douglas. 1994. "A Theory of Conformity." *Journal of Political Economy* 102 (5): 841–77.

Berinsky, Adam J., Gregory A. Huber, and Gabriel S. Lenz. 2012. "Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk." *Political Analysis* 20 (3): 351–68.

Bicchieri, Cristina, and Yoshitaka Fukui. 1999. "The Great Illusion: Ignorance, Informational Cascades, and the Persistence of Unpopular Norms." *Business Ethics Quarterly* 9 (1): 127–55.

Blair-Loy, Mary. 2003. *Competing Devotions: Career and Family among Women Executives.* Cambridge, MA: Harvard University Press.

Blau, Peter. M. 1964. *Exchange and Power in Social life.* New York: Wiley.

Bowles, Samuel, and Herbert Gintis. 2002. "Social Capital and Community Governance." *The Economic Journal* 112 (483): F419–36.

Buhrmester, Michael, Tracy Kwang, and Samuel D. Gosling. 2011. "Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data?" *Perspectives on Psychological Science* 6 (1): 3–5.

Centola, Damon, Robb Willer, and Michael Macy. 2005. "The Emperor's Dilemma: A Computational Model of Self-Enforcing Norms." *American Journal of Sociology* 110 (4): 1009–40.

Cialdini, Robert B. 1984. *Influence: Science and Practice.* Boston: Allyn and Bacon.

———. 2003. "Crafting Normative Messages to Protect the Environment." *Current Directions in Psychological Science* 12 (4): 105–9.

Cialdini, Robert B., and Trost, Melanie R. 1998. "Social Influence: Social Norms, Conformity, and Compliance." Gilbert, Daniel T., Fiske, Susan T., and Lindzey, Gardner (Ed). *The Handbook of Social Psychology.* p. 151-192. New York, NY, US: McGraw-Hill.

Cha, Christopher. 2015. "How to Survive a Korean Drinking Frenzy." *CNN*. Accessed December 27, 2016. http://www.cnn.com/2015/04/23/travel/parts-unknown-bourdain-korea-drinking/index.html.

Chwe, Michael Suk-Young. 2001. *Rational Ritual: Culture, Coordination, and Common Knowledge*. Princeton, N.J.; Woodstock: Princeton University Press.

Coleman, James S. 1990. *Foundations of Social Theory*. Cambridge, MA: Harvard University press, Belknap Press.

Collins, Lauren. 2016. "Are After-Work Drinks a Conspiracy Against Women?" *The New Yorker*, September 6. http://www.newyorker.com/culture/cultural-comment/are-after-work-drinks-a-conspiracy-against-women.

Cordes, Cynthia L., and Thomas W. Dougherty. 1993. "A Review and an Integration of Research on Job Burnout." *Academy of Management Review* 18 (4): 621–56.

Correll, Shelley J., Stephen Benard, and In Paik. 2007. "Getting a Job: Is There a Motherhood Penalty?" *American Journal of Sociology* 112 (5):1297–1339.

Correll, Shelley J., Cecilia L. Ridgeway, Ezra W. Zuckerman, Sharon Jank, Sara Jordan-Bloch, and Sandra Nakagawa. 2017. "It's the Conventional Thought That Counts: How Third-Order Inference Produces Status Advantage." *American Sociological Review* 82 (2): 297–327.

Coser, Lewis A. 1974. *Greedy Institutions; Patterns of Undivided Commitment*. Free Press.

Crawford, Vincent P., and Joel Sobel. 1982. "Strategic Information Transmission." *Econometrica* 50 (6): 1431–51.

David, Paul A. 1985. "Clio and the Economics of QWERTY." *American Economic Review* 75 (2): 332–37.

Deci, Edward L. 1971. "Effects of Externally Mediated Rewards on Intrinsic Motivation." *Journal of Personality and Social Psychology* 18 (1): 105–15.

De Mente, Boye Lafayette. 2014. *The Korean Way In Business: Understanding and Dealing with the South Koreans in Business*. Tokyo: Tuttle Publishing.

Derné, Steve. 1992. "Beyond Institutional and Impulsive Conceptions of Self: Family Structure and the Socially Anchored Real Self." *Ethos* 20 (3): 259–88.

Elster, Jon. 1989. "Social Norms and Economic Theory." *The Journal of Economic Perspectives* 3 (4): 99–117.

———. 1996. "Private Truths, Public Lies: The Social Consequences of Preference Falsification (Book)." *Acta Sociologica (Taylor & Francis Ltd)* 39 (1): 112–15.

Fincham, Robin. 1999. "The Consultant–Client Relationship: Critical Perspectives on the Management of Organizational Change." *Journal of Management Studies* 36 (3): 335–51.

Festinger, Leon. 1957. *A Theory of Cognitive Dissonance*. Stanford: Stanford University Press.

Frank, Robert H. 1988. *Passions Within Reason: The Strategic Role of the Emotions*. New York, NY: W. W. Norton & Company.

Gambetta, Diego. 2011. *Codes of the Underworld: How Criminals Communicate*. Princeton: Princeton University Press.

Gelman, Andrew, and Jennifer Hill. 2006. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. 1st edition. Cambridge University Press.

Glaser, Barney, and Anselm Strauss. 1967. *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Chicago: Aldine Transaction.

Gould, Roger V. 2002. "The Origins of Status Hierarchies: A Formal Theory and Empirical Test." *American Journal of Sociology* 107 (5):1143–78.

Granovetter, Mark. 1985. "Economic Action and Social Structure: The Problem of Embeddedness." *American Journal of Sociology* 91 (3): 481–510.

Hechter, Michael, and Karl-Dieter Opp. 2001. *Social Norms*. Russell Sage Foundation.

Henrich, Joseph. 2009. "The Evolution of Costly Displays, Cooperation and Religion: Credibility Enhancing Displays and Their Implications for Cultural Evolution." *Evolution and Human Behavior* 30 (4): 244–60.

Hirschman, Albert O. 1970. *Exit, Voice, and Loyalty: Responses to Decline in Firms, Organizations, and States*. Cambridge, MA: Harvard University Press.

Homans, G. C. 1961. *Social behavior: Its Elementary Forms*. New York: Harcourt, Brace & World.

Jerolmack, Colin, and Shamus Khan. 2014. "Talk Is Cheap: Ethnography and the Attitudinal Fallacy." *Sociological Methods & Research* 43 (2): 178–209. https://doi.org/10.1177/0049124114523396.

Kim, Minjae and Ezra W. Zuckerman Sivan. 2017. "Faking It is Hard to Do: Entrepreneurial Norm Enforcement and Suspicions of Deviance." *Sociological Science* 4: 580-610.

Kollock, Peter. 1994. "The Emergence of Exchange Structures: An Experimental Study of Uncertainty, Commitment, and Trust." *American Journal of Sociology* 100 (2): 313–45.

Kreps, David M, and Robert Wilson. 1982. "Reputation and Imperfect Information." *Journal of Economic Theory* 27 (2): 253–79.

Kunda, Gideon. 2009. *Engineering Culture: Control and Commitment in a High-Tech Corporation*. Philadelphia, PA: Temple University Press.

Kuran, Timur. 1995. *Private Truths, Public Lies: The Social Consequences of Preference Falsification*. Reprint edition. Cambridge, Mass.: Harvard University Press.

Landers, Renée M., James B. Rebitzer, and Lowell J. Taylor. 1996. "Rat Race Redux: Adverse Selection in the Determination of Work Hours in Law Firms." *The American Economic Review* 86 (3): 329–48.

Lawler, Edward E. 1986. *High-Involvement Management*. Jossey-Bass.

Lee, Robert. 2010. "Jingle All the Pay for Firms at Risk of Party Damages Claims." *The Korea Herald*. Accessed December 13, 2016. http://www.koreaherald.com/view.php?ud=20101213000892.

Mackie, Gerry. 1996. "Ending Footbinding and Infibulation: A Convention Account." *American Sociological Review* 61 (6): 999–1017.

Mason, Winter, and Siddharth Suri. 2012. "Conducting Behavioral Research on Amazon's Mechanical Turk." *Behavior Research Methods* 44 (1): 1–23.

Merton, Robert K. 1987. "Three Fragments From a Sociologist's Notebooks: Establishing the Phenomenon, Specified Ignorance, and Strategic Research Materials." *Annual Review of Sociology* 13 (1): 1–29.

Milgrom, Paul, and John Roberts. 1982. "Predation, Reputation, and Entry Deterrence." *Journal of Economic Theory* 27 (2): 280–312.

Miller, Dale T., and Cathy McFarland. 1987. "Pluralistic Ignorance: When Similarity Is Interpreted as Dissimilarity." *Journal of Personality and Social Psychology* 53 (2): 298–305.

———. 1991. "When Social Comparison Goes Awry: The Case of Pluralistic Ignorance." In *Social Comparison: Contemporary Theory and Research*, edited by J. Suls and T. A. Wills, 287–313. Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Miller, Dale T., and Deborah A. Prentice. 2016. "Changing Norms to Change Behavior." *Annual Review of Psychology* 67 (1): 339–61.

Munsch, Christin L., Cecilia L. Ridgeway, and Joan C. Williams. 2014. "Pluralistic Ignorance and the Flexibility Bias: Understanding and Mitigating Flextime and Flexplace Bias at Work." *Work and Occupations* 41 (1): 40–62.

Nunn, Nathan, and Raul Sanchez de la Sierra. 2017. "Why Being Wrong Can Be Right: Magical Warfare Technologies and the Persistence of False Beliefs." *American Economic Review* 107 (5): 582–87.

OECD Statistics. 2017. "Average Annual Hours Actually Worked per Worker." 2017. Accessed September 14, 2017. https://stats.oecd.org/Index.aspx?DataSetCode=ANHRS.

Parigi, Paolo, Jessica J. Santana, and Karen S. Cook. 2017. "Online Field Experiments: Studying Social Interactions in Context." *Social Psychology Quarterly* 80 (1): 1–19.

Perlow, Leslie A. 1997. *Finding Time: How Corporations, Individuals, and Families Can Benefit from New Work Practices*. Collection on Technology and Work. Ithaca, NY: Cornell University Press.

———. 1998. "Boundary Control: The Social Ordering of Work and Family Time in a High-Tech Corporation." *Administrative Science Quarterly* 43 (2): 328–57.

Phillips, Damon J., Catherine J. Turco, and Ezra W. Zuckerman. 2013. "Betrayal as Market Barrier: Identity-Based Limits to Diversification among High-Status Corporate Law Firms." *American Journal of Sociology* 118 (4): 1023–54.

Podolny, Joel M. and James N. Baron. 1997. "Resources and Relationships: Social Networks and Mobility in the Workplace." *American Sociological Review* 62:673–693.

Posner, Eric A. 2000. *Law and Social Norms*. Harvard University Press.

———. 2002. "The Signaling Model of Social Norms: Further Thoughts." *University of Richmond Law Review* 36 (2): 465-480.

Powell, Walter W., 1990. "Neither Market nor Hierarchy: Network Forms of Organizations." *Research in Organizational Behavior*, 12:295-336.

Power, Eleanor A. 2017. "Discerning Devotion: Testing the Signaling Theory of Religion." *Evolution and Human Behavior* 38 (1): 82–91.

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.

———. 1996. "Pluralistic Ignorance and the Perpetuation of Social Norms by Unwitting Actors." In *Advances in Experimental Social Psychology*, edited by Mark P. Zanna, 28:161–209. Academic Press. http://www.sciencedirect.com/science/article/pii/S0065260108602385.

Rainey, Carlisle. 2014. "Arguing for a Negligible Effect." *American Journal of Political Science* 58 (4): 1083–1091.

Rappaport, Roy A. 1999. *Ritual and Religion in the Making of Humanity*. Cambridge University Press.

Ridgeway, Cecilia L., and Shelley J. Correll. 2006. "Consensus and the Creation of Status Beliefs." *Social Forces* 85 (1): 431–53.

Rusbult, Caryl E. 1980. "Commitment and Satisfaction in Romantic Associations: A Test of the Investment Model." *Journal of Experimental Social Psychology* 16 (2): 172–86.

———. 1983. "A Longitudinal Test of the Investment Model: The Development (and Deterioration) of Satisfaction and Commitment in Heterosexual Involvement." *Journal of Personality and Social Psychology* 45(1): 101-117.

Schelling, Thomas C. 1981. *The Strategy of Conflict*. Reprint edition. Cambridge: Harvard University Press.

Sen, Amartya. [1977] 1979. "Rational Fools: A Critique of the Behavioural Foundations of Economic Theory." Pp. 87-109 in Frank Hahn and Martin Hollis eds., *Philosophy and Economic Theory*. Oxford: Oxford University Press.

Spence, Michael. 1973. "Job Market Signaling." *Quarterly Journal of Economics* 87 (3): 355–74.

———. 1981. "Signaling, Screening, and Information." *NBER*. p.319-358.

Swidler, Ann. 1986. "Culture in Action: Symbols and Strategies." *American Sociological Review* 51 (2): 273–86.

Trost, Jan E. 1986. "Statistically Nonrepresentative Stratified Sampling: A Sampling Technique for Qualitative Studies." *Qualitative Sociology* 9 (1): 54–57.

Turco, Catherine J. 2010. "Cultural Foundations of Tokenism: Evidence from the Leveraged Buyout Industry." *American Sociological Review* 75 (6): 894–913.

Uzzi, Brian. 1997. "Social Structure and Competition in Interfirm Networks: The Paradox of Embeddedness." *Administrative Science Quarterly* 42 (1): 35–67.

———. 1999. "Embeddedness in the Making of Financial Capital: How Social Relations and Networks Benefit Firms Seeking Financing." *American Sociological Review* 64 (4): 481–505.

Van Maanen, John. 1991. "Drinking Our Troubles Away: Managing Conflict in a British Police Agency." In *Hidden Conflict in Organizations*, edited by D. Kolb, J.M. Bartunek, 78-99. Newbury Park, CA: Sage Publications.

Van Maanen, J. and E. H. Schein. 1979. "Toward of Theory of Organizational Socialization." *Research in Organizational Behavior*, 1: 209-264.

Vaughan, Diane. 1986. *Uncoupling: Turning Points in Intimate Relationships*. Oxford University Press.

Wedeen, Lisa. 1999. *Ambiguities of Domination: Politics, Rhetoric, and Symbols in Contemporary Syria*. Chicago: University of Chicago Press.

Weeks, John R. 2004. *Unpopular Culture: The Ritual of Complaint in a British Bank*. Chicago: University Of Chicago Press.

Werr, Andreas, and Alexander Styhre. 2002. "Management Consultants - Friend or Foe? Understanding the Ambiguous Client-Consultant Relationship." *International Studies of Management & Organization* 32 (4): 43–66.

Willer, Robb. 2009. "Groups Reward Individual Sacrifice: The Status Solution to the Collective Action Problem." *American Sociological Review* 74 (1): 23–43.

Willer, Robb, Ko Kuwabara, and Michael W. Macy. 2009. "The False Enforcement of Unpopular Norms." *American Journal of Sociology* 115 (2): 451–90.

Williams, Joan C., Jennifer L. Berdahl, and Joseph A. Vandello. 2016. "Beyond Work-Life 'Integration.'" *Annual Review of Psychology* 67 (1): 515–39.

"Work net online survey" ctd in Han, Jong-soo. 2013. "Best Option for After-hour Gatherings is a Taste-tour." *News One*. Accessed December 27, 2016. http://news1.kr/articles/?1350860

Xiao, Zhixing and Anne S. Tsui. 2007. "When Brokers May Not Work: The Cultural Contingency of Social Capital in Chinese High-tech Firms." *Administrative Science Quarterly* 52:1–31.

Yamagishi, Toshio, Karen S. Cook, and Motoki Watabe. 1998. "Uncertainty, Trust, and Commitment Formation in the United States and Japan." *American Journal of Sociology* 104 (1): AJSv104p165-194.

Yamagishi, Toshio, and Midori Yamagishi. 1994. "Trust and Commitment in the United States and Japan." *Motivation and Emotion* 18 (2): 129–66.

**TABLES**

**Table 1**: Implications from existing approaches and this paper's proposition on when conformity to the norm will persist. The shaded cell indicates the condition under which the paper makes a different prediction from the existing studies, while the predictions are not contradictory with regard to persistence of the norm in the other conditions.

| | **When the norm is actually and visibly popular** <br><br>(i.e., when there is an accurate perception that the norm represents the group's values and preferences) | **When the norm is actually and visibly unpopular** <br><br>(i.e., when there is an accurate perception that the norm contradicts the group's values and preferences) |
|---|---|---|
| **When one's commitment to interactants is uncertain and needs to be demonstrated** | Typical individuals endorse the norm. As a result, conformity to the norm persists. <br><br>*When the norm actually is at odds with the group's values or preferences but appears to be endorsed, individuals privately do not like conforming to the norm but need to publicly appear like they endorse the norm. In such a situation, conformity to the norm persists. | **Prediction from existing studies:** <br><br>Typical individuals do not endorse the norm. Further, the norm is known to contradict the group members' values and preferences. Therefore, conformity to the norm stops. <br><br>**Prediction from this paper's theory:** <br><br>Typical conformity to the visibly unpopular norm signals that the conformist is willing to violate her own preference to meet the behavioral standard observed by others. Insofar as individuals seek this signaling value, conformity to the norm persists. |
| **When one's commitment to interactants has been demonstrated or does not need to be demonstrated** | Typical individuals endorse the norm. As a result, conformity to the norm persists. <br><br>*The norm might actually be at odds with the group's values or preferences even though it appears to be endorsed. But since individuals do not need to appear as if they also endorse the same values as others, conformity to the norm does not persist. | Typical individuals do not endorse the norm. They also do not need to signal commitment to their interactants, since it has already been demonstrated or does not need to be. Therefore, insofar as the norm contradicts the group's values or preferences, conformity to the norm does not persist. |

*Conditions identified by the literature as the case of "pluralistic ignorance"

**Table 2**: Study 1 Sample characteristics

| Characteristic | Client-facing role | Non-client-facing role | Total |
|---|---|---|---|
| Entry-level | 3 | 6 | **9** |
| Mid-career | 9 | 14 | **23** |
| Senior / Retired | 5 | 5 | **10** |
| **Total** | 17 | 25 | 42 |

| Characteristic | SMEs (including startups) | *Chaebols* | Foreign-owned /affiliated | Government / Gov't-owned | Total |
|---|---|---|---|---|---|
| Entry-level | 1 | 3 | 3 | 2 | **9** |
| Mid-career | 5 | 9 | 3 | 6 | **23** |
| Senior / Retired | 2 | 5 | 2 | 1 | **10** |
| **Total** | **8** | **17** | **8** | **9** | 42 |

**Table 3**: Study 2 Experimental Conditions

| *No Information* | **Dislike drinking (Prefer deviating from the norm)** | **Like drinking (Enjoys conforming to the norm)** |
|---|---|---|
| **Drink (Conform)** | *Suffering conformist* vs. *Baseline conformist* (n=53) | *Enthusiastic conformist* vs. *Baseline conformist* (n=43) |
| **Do not drink (Do not conform)** | *Nonconformist* vs. *Baseline conformist* (n=56) | -- |

| *Direct information* | **Dislike drinking (Prefer deviating from the norm)** | **Like drinking (Enjoys conforming to the norm)** |
|---|---|---|
| **Drink (Conform)** | *Suffering conformist* vs. *Baseline conformist* (n=124) | *Enthusiastic conformist* vs. *Baseline conformist* (n=110) |
| **Do not drink (Do not conform)** | *Nonconformist* vs. *Baseline conformist* (n=113) | -- |

Note: Direct information about the two candidates' capabilities and commitments is counterbalanced in each of *Direct information* conditions, meaning that in one set of condition a *Baseline conformist* is indicated to be more committed than a *Suffering conformist,* an *Enthusiastic conformist*, or a *Nonconformist*; and vice versa in another set of condition, which is why those conditions have twice as many subjects as do *No information* conditions.

**Table 4:** OLS regression models of the effect of being a *Suffering conformist* on (relative) preference as a relationship partner (Models 1, 4, and 5), on perceived commitment (Model 2), and perceived competence (Model 3).

| Dependent variable | Model 1<br>Preference as a relationship partner | Model 2<br>Perceived commitment | Model 3<br>Perceived competence | Model 4<br>Preference as a relationship partner | Model 5<br>Preference as a relationship partner |
|---|---|---|---|---|---|
| | b<br>(s.e.) | b<br>(s.e.) | b<br>(s.e.) | b<br>(s.e.) | b<br>(s.e.) |
| *Suffering conformist* | 1.08*<br>(0.50) | 0.88*<br>(0.35) | 0.24<br>(0.28) | 0.37<br>(0.42) | 0.32<br>(0.40) |
| Perceived commitment | | | | 0.81***<br>(0.10) | 0.76***<br>(0.09) |
| Perceived competence | | | | | 0.39***<br>(0.12) |
| Intercept | -1.14***<br>(0.29) | -0.88***<br>(0.21) | 0.15<br>(0.17) | -0.43<br>(0.26) | -0.88***<br>(0.21) |
| Adjusted R squared | 0.02 | 0.03 | <0.01 | 0.34 | 0.39 |
| N | 152 | 152 | 152 | 152 | 152 |

*p<0.05; **p<0.01; ***p<0.001; Note: Models 2 and 3 have different dependent variables from Models 1 4, and 5. Reference category in all models is the combined relative preferences for *Nonconformist* and *Enthusiastic conformist* from their respective conditions.

**FIGURES**

**Figure 1:** A picture from the corporate campaign where employees promote the "119" policy. The numbers stand for one drink, one venue, and a 9pm finish, which is stated in Korean in the bottom banner.



Photo from Lee (2010)

**Figure 2:** The number of informants who suggested that they drink more under certain conditions is indicated on the right-hand side of the above box. The below box indicates which informant (coded from 1 to 42) mentioned which condition as when they need to drink more. Note that three informants did not bring up any conditions under which they need to drink more, even though they did suggest that drinking in after-hour gatherings is nonetheless important.

| | | |
|---|---|---|
| | When the relationship is nascent | N=23 |
| | When facing (consequential) evaluations | N=21 |
| | When occupying a less favorable market position | N=17 |
| | When they want to stay in the relationship/group | N=12 |

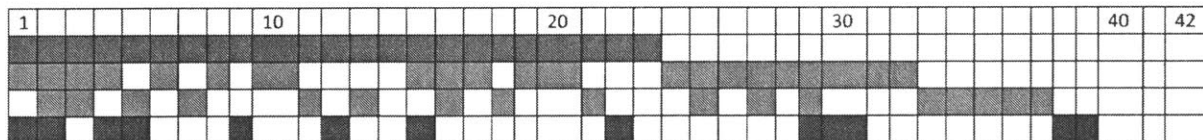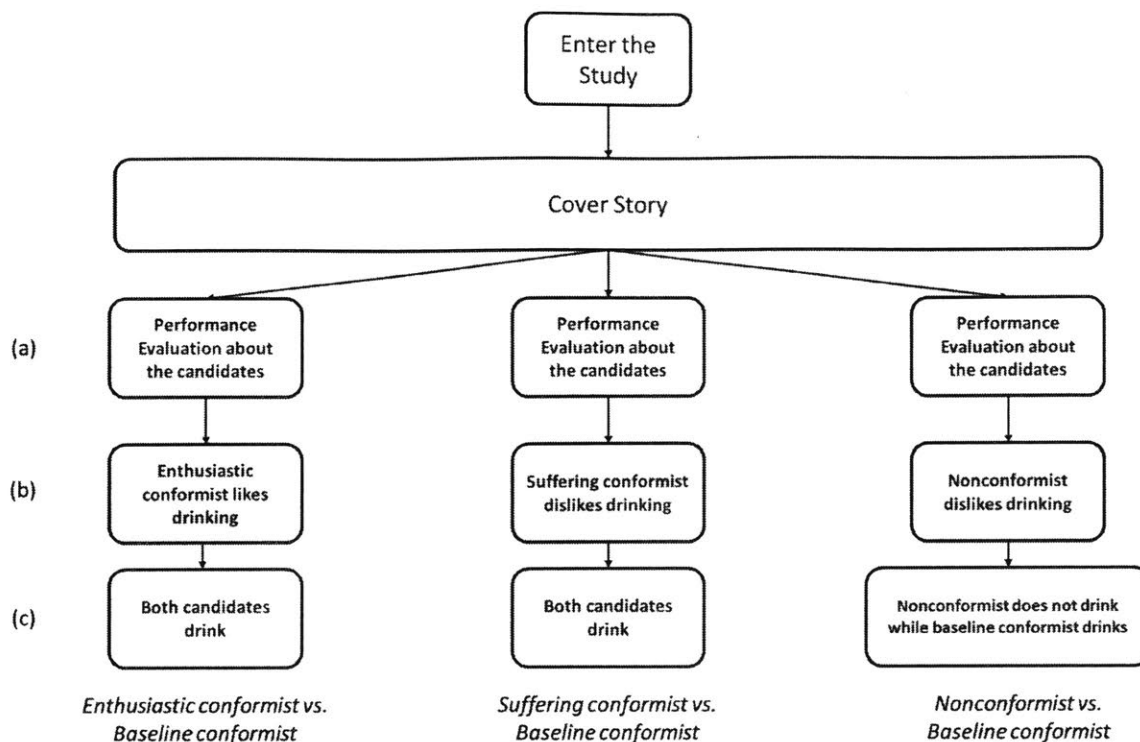| 1 | | | | | | | | | 10 | | | | | | | | | | 20 | | | | | | | | | | 30 | | | | | | | | | | 40 | | 42 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|

**Figure 3:** Overview and Flow of Study 2



Note: Subjects are randomly distributed to one of the three sets of conditions. Step (a) was shown only to subjects in *Direct Information* conditions; for subjects in the *No information* conditions, they were taken straight to step (b) after reading about the setting. In step (b), subjects were explicitly told the preferences of *Enthusiastic conformist, Suffering conformist* or *Nonconformist*, while they were not explicitly told anything about the preference of *Baseline conformist*. In step (c), subjects were told whether the candidates drank or not.

**Figure 4**: Results for within-subject comparisons and between-subject comparisons. Within-subject comparisons are t-tests for the conditions beneath where indicated. Between-subject comparisons are t-tests between relative preferences for *Baseline conformists* of two indicated conditions. The between-subject comparisons are therefore equivalent to difference-in-difference tests. Subjects in each condition evaluated two actors. Subjects were asked to evaluate the statements "Associate Kim chose to collaborate with [*Baseline conformist*]'s team" and "Associate Kim chose to collaborate with [*Nonconformist, Suffering conformist,* or *Enthusiastic conformist*]'s team" on scales ranging from 1 (Very unlikely) to 10 (Very likely). All results are from *No information* conditions, where there is no direct information about protagonists' commitment.
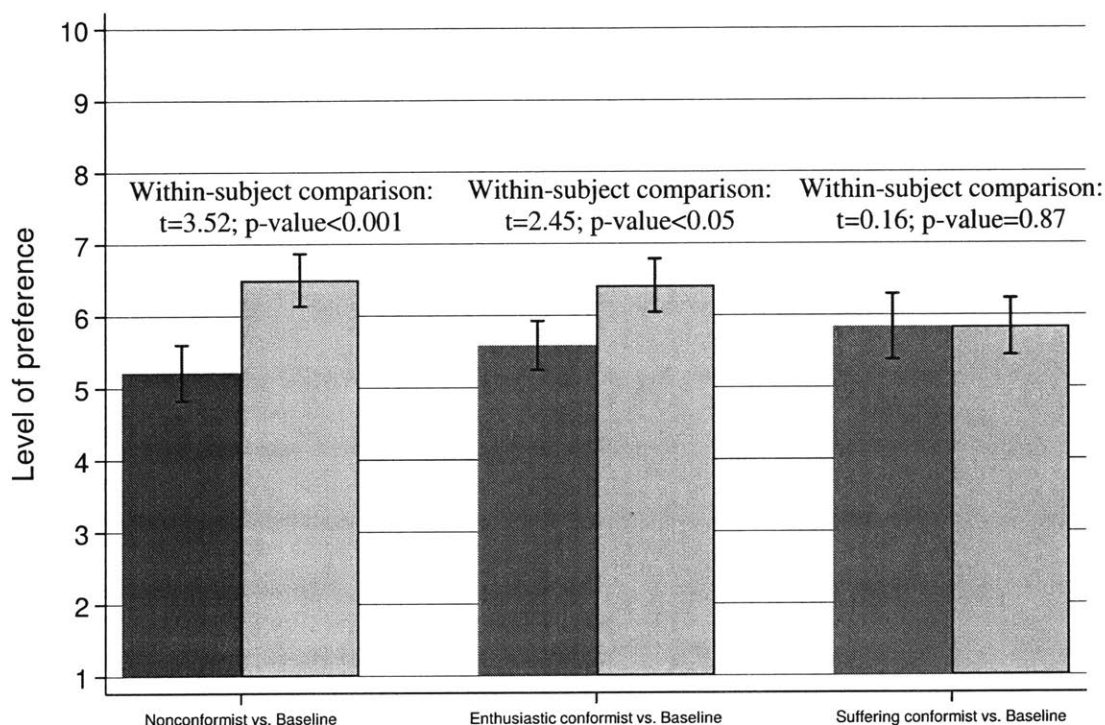


Note: To make sure that the pair-wise comparison is not a byproduct of statistical chance, I also used two-way ANOVA as a global test, where I compare the perceived relative preference for a *Suffering conformist* to that for a *Nonconformist* and that for an *Enthusiastic conformist*. And this analysis validates that preference for a *Suffering conformist* is greater than that for the other two categories (F[1,150]=4.76, P<0.05), validating that a *Suffering conformist* is meaningfully (perceived to be) preferred over the other categories.

[a]Between-subject comparison between the relative preference for the *Suffering conformist* and that for the *Nonconformist*: t=2.24; p-value<0.05

[b]Between-subject comparison between the relative preference for the *Enthusiastic conformist* and that for the *Nonconformist*: t=0.53; p-value=0.60

**Figure 5:** Results for direct (within-subject) comparisons and indirect (between-subject) comparisons. Within-subject comparisons are t-tests for the conditions beneath where indicated. Between-subject comparisons are t-tests between relative preferences for *Baseline conformists* of two neighboring conditions in the figure. Subjects in each condition evaluated two actors. Subjects were asked to evaluate the statements "Associate Kim chose to collaborate with [*Baseline conformist*]'s team" and "Associate Kim chose to collaborate with [*Nonconformist, Suffering conformist, or Enthusiastic conformist*]'s team" on scales ranging from 1 (Very unlikely) to 10 (Very likely). All results are from *Direct information* conditions, where commitment of a *Nonconformist, Suffering conformist,* or *Enthusiastic conformist* is demonstrated to be higher than that of a *Baseline conformist*.
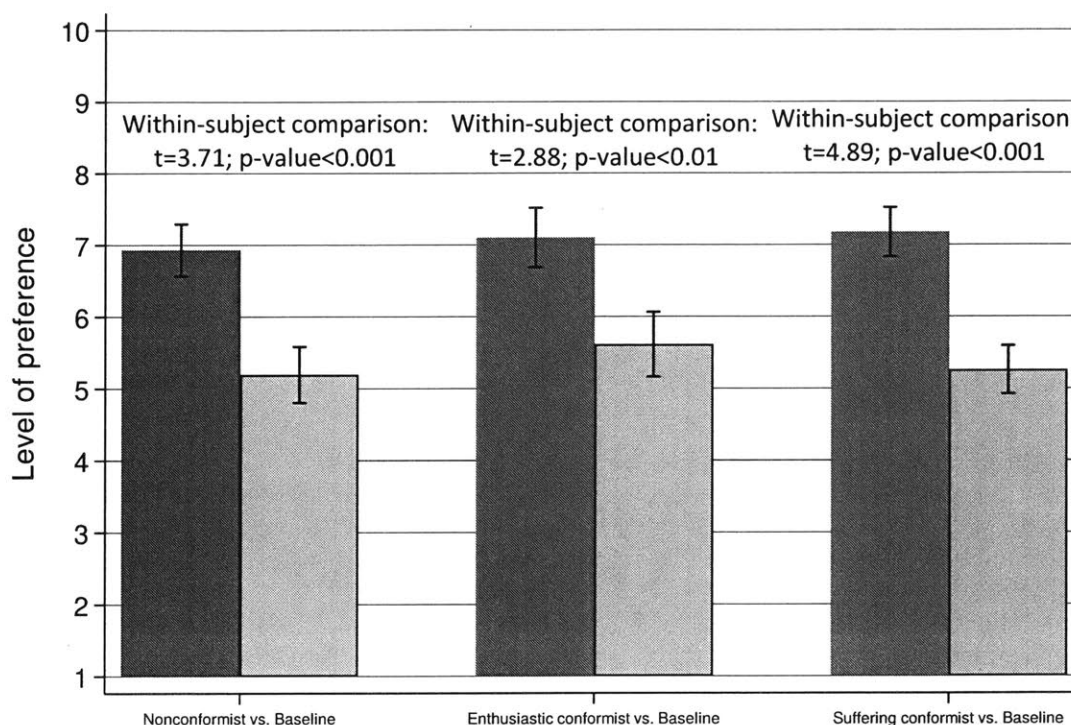


Note: I here also use a two-way ANOVA as a global test where I compare the relative preference for a *Suffering conformist* to the relative preference for a *Nonconformist* and that for an *Enthusiastic conformist*. The test shows that in the first set of *Direct information* conditions (where commitment of a *Nonconformist, Suffering conformist,* and *Enthusiastic conformist* is directly shown to be greater), there is no evidence that a *Suffering conformist* is preferred over a *Nonconformist* and an *Enthusiastic conformist* ($F[1,166]=0.30$, $P=0.59$), in contrast to the significant effect in *No information* conditions.
[a]Between-subject comparison between the relative preference for the *Suffering conformist* and that for the *Nonconformist*: $t=30$; p-value=0.77

# Chapter 2

Faking It Is Hard to Do:
Entrepreneurial Norm Enforcement and Suspicions of Deviance
(coauthored with Ezra W. Zuckerman Sivan)

'Doesn't your position as a delegate to the Political Consultative Conference give you
some protection?' I asked my friend.
'I hear the Maoists want to abolish that organization. *They call it a collection of
radishes, red on the outside but white inside.* They claim that while all the
delegates talked as if they support the Communist Party, in actual fact they
oppose the party,' she said. (Cheng 1986:73–74, emphasis added)

## INTRODUCTION

One of the foundational ideas of sociology is that norms are a primary source of social action and

order (Coleman 1990; Homans 1950; Parsons 1937). But an enduring question is why members

of a group might expend effort to ensure that others conform. Rational choice theorists have

observed that "enforcement" (i.e., sanctioning norm violators and/or encouraging others to

conform) is a public good and that the rational strategy is to let others bear the cost of providing

this public good (Head 1974; Oliver 1980; Olson 1965:2; Rowley and Peacock 1975; Samuelson

1954). But more recent research suggests that individual actors might gain private utility (i.e.,

"selective incentives"; Olson 1965) from enforcement (Adut 2004, 2012; Becker 1963; Kuran

1995; Sunstein 1996:910–14). In particular, Willer, Kuwabara, and Macy (2009; cf. Centola,

Willer, and Macy [2005] and Jordan et al. [2016]) argue that audiences regard an enforcer as

more sincere in her compliance than those who do not enforce even though the enforcer is

actually more likely to have privately deviated from the norm, and that the "illusion of sincerity"

available from enforcement explains why individuals are often motivated to enforce. The larger

implications of this argument are twofold. First, an enforcer will appear more sincere in her

compliance than a "bystander" who is not incurring a personal cost to enforce. Second, norms

will be enforced as long as groups have insecure members who have reason to portray their conformity as sincere.

A puzzle emerges from a consideration of this social logic, however. Insofar as there might be ulterior motives behind norm enforcement, and insofar as audiences may themselves engage in enforcement to mask deviance in other social situations, it is odd that audiences do not recognize possibilities for such ulterior motives and discount for them. In fact, various empirical accounts suggest that rather than enhancing a perception of sincerity, enforcement often does raise audience suspicion that the enforcer is masking deviance. Consider research on purges such as the Chinese Cultural Revolution (Walder 2006, 2009; also, see Goode and Ben-Yehuda [1994]), whereby patterns of cross-accusation of deviance between different factions are documented. As suggested by the epigraphical quotation, such enforcement often raised suspicions that avowals of fealty to Mao were driven by the ulterior motive of wanting to protect themselves from allegations of "revisionism." Similarly, when contemporary firms engage in social activism to fend off suspicions of unethical practices raised by social movement activists (Ingram, Yue, and Rao 2010; McDonnell and King 2013), enforcement often elicits doubts about their commitment and true motives (Yoon, Gurhan-Canli, and Schwarz 2006; cf. Carlos and Lewis 2017). The implication of these accounts then runs counter to the one summarized above: norms might in fact be under-enforced because individuals are wise to such risk of enforcement. So which is it? Why and when might enforcement elicit suspicion of deviance while it at other times successfully creates an illusion of sincerity?

We ground our approach to this question in an analysis of a general problem faced by social observers—that of inferring whether an interactant's motives are sincere or instead she is driven by an ulterior motive. The "sincere motive" inference often seems plausible because

observers know that the enforcer is incurring a personal cost to help maintain the norm (Becker 1960; Schelling 1980). But when audiences recognize that there might be possible strategic reasons to feign commitment—in situations we label "high accusability"—they are wise to the possibility that an individual might try to avoid accusations of deviance by engaging in norm enforcement. We argue that to solve the inferential contest between the "sincere motive" and "ulterior motive" that arises in such contexts, audiences look for situational cues. In particular, social situations vary in the degree to which actors are charged with responding to normative deviance. At one end of the continuum, they might be mandated to to assess one another's commitment to norms (e.g., Jordan et al. 2016; Willer et al. 2009). Someone who is given such a mandate cannot claim to be unaware of the deviance—that is, she does not enjoy "plausible deniability." As such, to fail to enforce the norm is to undermine it. By contrast, when there is no mandate and thus enforcement requires an interruption of social interaction—that is, one must take "entrepreneurial" initiative (Becker 1963; Sunstein 1996)—the alternative to enforcement is to do nothing and remain a bystander. Such a bystander enjoys plausible deniability of the deviance in that she can credibly claim that she either did not observe the deviance or did not appreciate its significance. As such, it is natural to question why someone would step out of the bystander role and go out of her way to punish someone: the possibility that she is masking deviance thus emerges as highly salient.

After developing our theoretical framework, we test it using a series of online vignette experiments. In the experiments, subjects read about a fictitious group of college students who express their opinions on alcohol use on a college campus. In these studies, we demonstrate that audiences perceive an enforcer to be sincere when there is a mandate to respond to, but audience suspicions are aroused when (1) audiences see possible strategic motives for appearing

committed (i.e., in a situation of high accusability) and (2) an enforcer takes entrepreneurial

initiative to enforce. We conclude by noting several implications of our theory and findings.

Most notably, whereas recent research sees the pursuit of reputational benefits as

straightforwardly stimulating prosocial behavior (e.g., Jordan et al. 2016; Willer 2009; see

Simpson and Willer [2015] for review), our analysis indicates that such benefits (and thus the

prosocial behavior) may be limited or enabled by common situational cues and audience

perceptions of strategic motive. Our analysis also implies that potentially destabilizing

undercurrents of suspicion might lurk even in many social settings where norms are superficially

defended.


## THEORY

*Deepening the Puzzle*

Willer et al. (2009) offer two possible explanations for why enforcement creates an

illusion of sincerity despite rational suspicions that the enforcer might be trying to hide her own

deviance.[21] One idea comes from research that suggests that human beings often err by

projecting their own motivations onto others (Miller and McFarland 1991; Prentice and Miller

1993). The other possible explanation is "correspondence bias," a cognitive heuristic by which

actors assume that others' public behaviors accurately reflect their private beliefs (Gilbert and

Malone 1995; Jones and Harris 1967). These explanations essentially suggest that cognitive

---

[21] Willer et al. (2009) focus on the case of a norm from which individuals privately dissent (an "unpopular norm"). But insofar as the norm is perceived to be privately endorsed by individuals who publicly conform (which is the case in Willer et al. 2009), audiences have the same inferential problem as they would have with enforcement of a norm that is actually privately endorsed by members of the group because they, in either case, think that the norm (whether popular or unpopular) is privately endorsed by other individuals (Prentice and Miller 1993; cf. Kim 2017). We discuss this issue further in the Methodological Appendix in the online supplement.

difficulties prevent audience members from taking the enforcer's perspective even though the shoe is often on the other foot.[22]

The experimental evidence presented by Willer et al. (2009) is consistent with this suggestion. In their experiment, subjects were informed of a situation where three actors (1) were asked to privately evaluate a purportedly high-quality but actually nonsensical text,(2) had a public discussion in which one turned out to be a deviant and two turned out to be conformists, and (3) were asked to publicly evaluate their fellow group members. Subjects were then told of one (randomly selected) conformist's evaluation of fellow group members, either in a condition in which the conformist enforces the norm by criticizing the deviant or in a condition in which the conformist does not enforce the norm by evaluating all fellow group members equally. Subjects were more likely to think that the enforcer was more sincere in his compliance than the non-enforcer, which suggests that enforcement can occur without eliciting suspicions of ulterior motives.[23] The larger implication is that many norms might be enforced even if there is significant private dissent (see Study 1 and 2 in Willer et al. [2009]).

But there is reason to question how enforcers can appear more sincere in their compliance with the norm than do bystanders (i.e., those who do not enforce). Put differently, even if audiences face cognitive challenges in recognizing ulterior motives in others, it seems that these challenges are often overcome. Accordingly, numerous empirical accounts document instances where (1) sincerity in compliance with the norm is considered important in maintaining one's

---

[22] If it is situationally impossible for the enforcer to be the deviant (e.g., victims of sexual assault disputes in which victims cannot be the perpetrator of the same sexual assault case), enforcement may not attract suspicions of deviance. Even in such cases, however, the enforcer may encounter a reputational backlash (1) because they appear to profit from public enforcement or (2) because the publicity of their enforcement is distasteful (Adut 2008).

[23] See the Methodological Appendix in the online supplement for more details.

status in the group (Willer 2009) and (2) norm enforcers are viewed with suspicion that they themselves broke the norm. Notable examples include the following cases.

*The Chinese Cultural Revolution (see especially Walder [2009]).* In this well-known political purge, Mao declared that there were hidden "revisionists" and demanded self-regulation. In the tumult that followed, high school and university students in Beijing (then spreading to the rest of the country) began accusing party officials and one another of being revisionists. Such accusations fueled a general environment of suspicion (Walder 2006).

*Socially responsible activities by corporations.* Firms in the contemporary market often engage in socially responsible activities through which they encourage others to comply with a norm, but such activities often elicit suspicion that they are engaging in those activities in order to cover up their own deviant deeds (Yoon et al. 2006; cf. Carlos and Lewis 2017).[24]

*Homophobia.* Individuals who publicly ridicule and attack gay individuals often elicit suspicion that they are overcompensating for their insecurity in their masculinity (cf. Humphreys 1970; Willer et al. 2013). The very term "homophobia" implies that bias against lesbian, gay, bisexual, and transgender (LGBT) individuals is often driven by the attacker's insecurities and fears.[25]

These examples suggest that audience members often view norm enforcers with more suspicions of deviance than they view bystanders who do not enforce.[26] Our challenge then is to

---

[24] Such terms as "greenwashing" (encouraging the adoption of environment-friendly policies and activities to cover their own environmental misdeeds) and "pinkwashing" (advertising and encouraging the adoption of LGBT-friendly policies to cover their own misdeeds) reflect such suspicion.

[25] Perhaps the more illustrative examples are various sayings in many cultures that imply a recognition of insecurity that stimulates norm enforcement. One of many English examples is, "whoever smelt it dealt it." Similar sayings abound in other cultures: A Chinese example is, "one who yells out 'catch the thief' is the thief"; a Korean version is, "the dog that scolds another dog for being dirty is actually dirtier"; and a Hebrew version is, "all who accuse others of a disqualification suffer from that very disqualification."

[26] At the same time, these accounts document that enforcement occurred despite the fact that it breeds suspicion (cf. McDonnell 2015). We further discuss this point in the General Discussion section.

develop a theory that addresses why and when acts of norm enforcement elicit suspicions of deviance in some cases, whereas in other situations, audiences deem enforcers to be sincere in their compliance.

*Theoretical Framework*

Our theory focuses on the two possible motives that might be inferred from any act of enforcement (cf. Hahl and Zuckerman 2014).[27] One possible inference is that the enforcer is highly committed to the norm and is motivated to protect it: the fact that the enforcer is apparently willing to bear a personal cost in order to contribute to the public good bolsters the impression of commitment. We will call this the "sincere motive" inference. In the instance of the Chinese Cultural Revolution, one who accuses another of being a revisionist might elicit less suspicion she is a closeted revisionist because she is willing to speak up and bear the cost of having to navigate the social situation. But the other possible inference that audiences can draw is that enforcement is motivated by the desire to fabricate the impression of commitment to the norm. The key assumption here is that audience members are aware that the actor has an incentive to bolster the impression that she is sincere in her conformity. Given the general

---

[27] Research on prosocial behavior and costly punishment in evolutionary psychology and behavioral economics is broadly relevant here. The question in those lines of work is whether actors are perceived as prosocial when they actively sanction others who fail to engage in prosocial behavior (e.g., if they fail to allocate money to others in a dictator game). Such sanctioning can be motivated either by sincere commitment to the group or by vengeance against the norm violator (Brañas-Garza et al. 2014; Fehr and Gächter 2002; Henrich et al. 2001; Raihani and Bshary 2015; cf. Jordan et al. 2016). A key implication is that enforcers are more likely to be perceived as committed when they are not the primary victim of deviance and thus less likely to be motivated by vengeance (Feldman Hall et al. 2014; Raihani and Bshary 2015:101). This implication is consistent with the approach taken here in that the key question is under what situational conditions the interpretation of ulterior motive is salient. However, this line of work does not speak directly to the contexts we discuss in our article because this work focuses on cases where there is no debate about what is and is not prosocial behavior (cf. Herrmann, Thöni, and Gächter 2008). Therefore, there is no way for these evolutionary models to account for the possibility of (hidden) deviance from a norm motivated by principled objection to that norm on the basis of the endorsement of an alternative norm. Our framework encompasses such a possibility.

awareness of the possibility of such a strategy, an enforcer might seem like a closeted deviant who has an ulterior motive to mask deviance. We will call this the "ulterior motive" inference. Lastly, note that even if audiences suspect an ulterior motive, this does not mean they are certain of the enforcer's lack of commitment. This distinction will become important when we discuss the implications of suspicion elicited by enforcement.

*(1) Situations of High Accusability Versus Situations of Low Accusability*

Let us now clarify the contextual conditions under which audiences are most likely to face a contest between sincere and ulterior motive inferences. The most general condition is one in which any participant in the situation has a reasonable fear of being credibly accused—what we term as a situation of "high accusability." Such a situation has three basic features: (a) deviance is known to have occurred to the point that there is reason to fear that social support for the norm will unravel, (b) it is widely believed that there are additional deviants whose identities are unknown, and (c) credible accusations of closeted deviance are likely. Whenever condition (a) is in place, someone who sincerely believes in the norm has an incentive to sanction a public deviant and thereby bolster the norm. And if it is either not common knowledge that there are closeted deviants (condition b) or there is little basis for thinking that closeted deviance might face a credible accusation (condition c), there is no reason to suspect such a norm enforcer of having an ulterior motive because closeted deviants have no reason to fear accusation (cf. Horne 2001, 2007). But insofar as conditions (b) and (c) are both in place, thus creating what we call a situation of high accusability, the sincere motive inference competes with the ulterior motive inference in the minds of an audience.

The cases reviewed above illustrate common examples of high-accusability situations. For instance, contemporary firms are increasingly exposed to the threat of being branded as ethically compromised by social movement activists (e.g., King and Pearce 2010). Situations of moral panic (Goode and Ben-Yehuda 1994) and purges often feature high accusability as well. It is also important to highlight that in a situation of high accusability, the fear of being accused can spread to all actors, including truly sincere conformists: even if one is free of deviance, allegations might nevertheless stick if there is no definitive proof of innocence. This engenders two key features of high-accusability situations as contrasted with low-accusability situations: (i) enforcement is an attractive strategy for actors (cf. Centola et al. 2005:1031), and (ii) audiences face an inferential contest (because a closeted deviant might be dissimulating to avoid credible accusations).

*(2) Entrepreneurial Enforcement Versus Enforcement in Response to a Mandate*

Yet some evidence suggests that enforcement might still successfully deflect suspicion of closeted deviance even in high-accusability situations (Adut 2004; Jordan et al. 2016; Reilly 2016; Willer et al., 2009). Thus, although high accusability may be a necessary condition for suspicion to be aroused, it is not sufficient. Hence, we identify a second condition (and another quite general context of enforcement) that is necessary for audience suspicion to be aroused— i.e., entrepreneurial enforcement. Situations (and roles within them) vary in the extent to which someone who did not enforce—a bystander—would enjoy plausible deniability with regard to the existence and/or significance of the deviance. At one end of the continuum are situations that mandate actors to express their views on the norm violation. The situations presented in existing experimental studies (see especially Simpson, Harrell, and Willer [2013] and Willer et al. [2009];

also, see Jordan et al. [2016] for work in evolutionary psychology) tend to induce enforcement via a mandate. In these situations, individuals are specifically asked what they think of others and thus effectively asked whether they are willing to uphold group norms or not. In such a context, to refrain from enforcing a norm or standard is effectively to endorse deviance or poor performance. As such, even someone who does not have an ulterior motive to enforce would have no choice but to enforce, lest she endorse the deviance. This implies that the audience will have no reason to infer anything but a sincere motive.

However, situations that mandate actors to express their views on the norm violation seem relatively rare. More typical cases of norm enforcement seem to be those in which the choice is either to stand by and let the flow of social interaction continue or to interrupt this flow and call out deviance. An enforcer in such a context is a "moral entrepreneur," in Becker's (1963) classic phrasing, or a "norm entrepreneur," in the more recent coinage by Sunstein (1996; e.g., Adut 2004, 2005, 2008; Reilly 2016; cf. Fine 1996).[28] The question of motive is highly salient for such an entrepreneur in a way that it is not when there is a mandate. In particular, were one to remain a bystander, one could plausibly claim that one did not witness the deviance or did not recognize its significance. A bystander thus does not seem to endorse the deviance. Conversely, someone who does take action should elicit suspicion: what motivated her to sanction the deviant when she did not have to say anything? Accordingly, when an enforcer takes entrepreneurial initiative to publicly identify wrongdoing, the voluntary nature of her enforcement serves to raise suspicion that the enforcer is driven by the ulterior motive of trying

---

[28] Sunstein (1996) suggests that norm entrepreneurs are those who enforce new norms and thus are often "interested in changing social norms." However, because many norms are rarely enforced (and thus are "fragile"), norm entrepreneurship can be more broadly defined as the enforcement of norms that are rarely enforced or not explicitly established (see his discussion on p. 909). As elaborated on previously, our empirical tests examine the effect of the enforcement of such a fragile norm.

to cover up her own deviance. But if she were to remain silent then she would be able to plausibly claim that she did not know about norm violation or that she did not think the norm was threatened. As such, entrepreneurial enforcement should raise suspicion under situations of high accusability. Thus, as depicted in Table 1, we predict that when the situation is both one of high accusability and there is no mandate for enforcement, an enforcer will attract suspicion; otherwise, the enforcer will seem sincere. This theoretical framework is summarized in the following proposition:

*Proposition: Insofar as a social situation is one where (1) closeted deviants have a reasonable fear of being credibly accused and (2) norm enforcement occurs via entrepreneurial initiative on the part of enforcers, enforcement elicits doubts about the enforcer's commitments. Conversely, when either of these conditions is absent, enforcement will signal commitment to the norm.*

In sum, our theory explains why enforcement can project an illusion of sincerity (Willer et al. 2009; cf., Jordan et al. [2016]) but also why it more often elicits suspicion. In particular, whereas insecure actors might wish to hide behind the illusion of sincerity under situations of high accusability, this seems realistic only in (seemingly rare) cases where actors are given a mandate to enforce the norm but not where enforcement requires entrepreneurial action. Consequently, the suspicion bred by enforcement might generally result in under-enforcement of norms, whereby actors refrain from enforcing lest they appear suspicious. We elaborate on this implication of our theory (and notable exceptions to this implication) in the General Discussion section below.

**EMPIRICAL SETTING AND RESULTS**

We provide empirical validation for our proposition with online experiments. Subjects for each of our studies were recruited through Amazon Mechanical Turk and limited to those with American internet protocol addresses.[29] This online experimental setting is useful because we were looking for subjects whose characteristics reflected the general (American) population instead of an audience that has specific knowledge or skills (Berinsky, Huber, and Lenz 2012; Buhrmester, Kwang, and Gosling 2011). However, subjects may pay less attention when participating in online experiments compared to those who are in laboratory settings in universities. As has become standard, we addressed this challenge in two ways. First, we broke down our experiments into different steps. The idea here is for subjects to receive a small amount of information at each click instead of reading too much information at once. Second, we asked them a series of attention questions to track whether they were understanding our instructions correctly (Mason and Suri 2011).

*Studies 1 and 2: Demonstrating Suspicions of Deviance Cast upon Norm Entrepreneurship*

Studies 1 and 2 are designed to test three predictions applied to situations in which there is a publicly endorsed norm (i.e., there is no direct information on whether it is privately endorsed or disapproved of; cf. Prentice and Miller 1993; Kim 2017; Willer et al. 2009). The predictions emerge from the two dimensions captured in our proposition and Table 1, reflecting the two contextual conditions that animate our theory: (1) whether enforcement is entrepreneurial or mandated and (2) whether the situation is one that is of high versus low accusability. The first

---

[29] We limited the subject pool to Amazon Mechanical Turk participants who are located in the United States, experienced in the platform (i.e., completed more than 100 online surveys in the platform before), and reliable (i.e., had their previous surveys approved—as opposed to unapproved for omission or inattention—by surveyors more than 90 percent of the time).

prediction is a baseline prediction that derives from study 3 of Willer et al. (2009), whereby the illusion of sincerity is created by a norm enforcer. Our theory includes this prediction as well, though it explicitly limits it to situations in which enforcement occurs in response to a mandate.[30] Thus, the objective of the first prediction is to see how someone who enforces in response to a mandate appears as compared to someone who does not enforce in response to a mandate.

*Hypothesis 1: An enforcer responding to a mandate is more likely to be seen as sincerely endorsing the norm than someone who is given the same mandate but does not enforce.*

By contrast, when high accusability and the absence of a mandate raise the salience of the ulterior motive inference, an entrepreneurial enforcer enjoys less plausible deniability that she did not know about or appreciate the norm violation. Entrepreneurial enforcement is thus suspect.

*Hypothesis 2: In a situation where accusability is high, an enforcer taking entrepreneurial initiative is less likely to be seen as sincerely endorsing the norm than a bystander.*

Finally, we test whether a situation of high accusability (and not just entrepreneurial initiative) is a necessary condition to make the ulterior motive inference salient.

---

[30] The Methodological Appendix in the online supplement details how our studies compare with study 3 of Willer et al. (2009), who suggested that enforcement always elicits a higher perception of commitment than that of a bystander.

*Hypothesis 3: Entrepreneurial enforcers are more likely to be viewed as sincerely endorsing the norm in a situation of low accusability than in a situation of high accusability.*

In order to test these hypotheses, we designed two studies. In study 1, we simulated a situation of high accusability and tested hypotheses 1 and 2. In study 2, we compared enforcement in a situation of high accusability to that in a situation of low accusability and thereby tested hypothesis 3. In addition to confirming that the suspicions aroused by entrepreneurial enforcement indeed derive from inferences of strategic motive (hypothesis 3), study 2 plays another important role by testing whether interrupting the flow of social interaction in and of itself raises suspicion of deviance. Our focus in these studies was on distinguishing between norm entrepreneurship, at one end of the continuum, and enforcement in response to a mandate, at the other end. One might argue that an entrepreneurial enforcer is suspected of deviance not because audiences inferred an ulterior motive for enforcement but because audiences suspect anyone who interrupts the flow of social interaction in a situation when normative deviance is known to have occurred. But insofar as entrepreneurial enforcement does not raise suspicion in situations of low accusability (study 2), this implies that interruption in situations of deviance does not necessarily raise significant suspicion.

**Study 1 Design and Procedures**

*Recruitment and Conditions*

Five hundred and nine United States–based subjects[31] were recruited via Amazon

Mechanical Turk in November and December of 2015. Each subject was paid $0.75 for

participating. After consenting to participate in the experiment, subjects were told that they

would participate in a "social perception task" that was designed to "assess [their] ability to

accurately infer the responses of individuals from a group of college freshmen." Subjects were

further told that "[t]his group of freshmen recently participated in an orientation session at their

college and submitted their responses to an anonymous questionnaire."

Subjects were randomly assigned to one of four sets of high-accusability conditions—

"*High-Accusability/Mandated and Enforce*," "*High-Accusability/Mandated and Not Enforce*,"

"*High-Accusability/Entrepreneurial*," and "*High-Accusability/Baseline*." Subjects, therefore,

were not aware of what was happening in other conditions (or, for that matter, that there were

other conditions). The level of accusability indicates the extent to which the threat of accusation

looms over actors in the situation. The *Mandated* conditions (*Mandated and Enforce* and

*Mandated and Not Enforce*) presented a setup whereby the fictive group members were

specifically asked to assess a public deviant (but we told subjects only about the assessment of

the first group member who was asked to assess the public deviant). In the *Not Mandated*

conditions (i.e., *Entrepreneurial* conditions and *Baseline* condition), no such mandate was given,

and enforcement required initiative. In the *Entrepreneurial* conditions, we told subjects about the

---

[31] In all studies in this article, all subjects who completed the experiment were included in the analysis.
Responses to attention checks (seven or eight attention questions total) show that subjects understood the
vignette generally well (more than 90 percent of the subjects answered all attention questions correctly).
Analysis without subjects who failed one or more attention checks provides substantively the same
results.

assessment by the (one and only) entrepreneurial enforcer on the public deviant. And in the *Baseline* condition, neither was the mandate given nor did the enforcement take place. These conditions are summarized in Table 2.

Finally, note that each of the conditions besides the *Baseline* condition has three sub-conditions based on the specific manner by which enforcers invoke or do not invoke normative principles. This is more exploratory in nature but is important because we have no baseline for knowing how various enforcement activities are perceived, at least in this experimental paradigm. We elaborate on this issue below.

*Study 1 Vignette*

We first lay out the *Baseline* condition, in which neither mandate nor enforcement takes place. After doing so, we describe how other conditions in study 1 are different.

*High-Accusability/Baseline condition.* After the introduction, subjects were told the following story in four consecutive screens. First, they learned that (1) in the prior month, a new cohort of college freshmen arrived at a college campus in the United States and that (2) six of these students who "had met each other only briefly before" and were about to spend the entire year in the same dorm (labeled "students 1 through 6") had participated in an orientation session. Subjects were then given details about that orientation session and were told that they would be asked to evaluate the students' behavior. In particular, subjects learned that the six students had participated in a "Bloomsbury Ethics Roundtable," an academic tradition in the philosophy of ethics in which participants express their opinions on an ethical question, and that this process had two stages—first, writing their opinions in a private, anonymous questionnaire, and second,

having a public discussion.[32] Subjects learned that the students were informed about these stages in advance. In particular, the six students were asked to fill out an anonymous questionnaire on "whether or not the college should ban alcohol on its campus." Subjects also learned that after all the students had filled out their anonymous questionnaires, student 2 asked the administrator "why [the administrator] asked [the students] to fill out those questionnaires" when they "have to give their opinions publicly anyway." The administrator replied by saying that "some students [in previous years] were influenced by other students when they debated the issue publicly" and that the administrator is "interested in whether the discussion influences students' opinions." At this point, subjects were asked some attention-checking questions.[33]

Afterward, subjects learned about stage 2 in three consecutive screens. First, subjects learned that students 2, 4, 1, 3, and 6 had been called upon (in allegedly random order) to express their opinions, and each of them said that the college should not ban alcohol. Next, subjects learned that before the administrator could call the last student (student 5), "an accident" occurred. In particular, the administrator who was collecting the anonymous questionnaires dropped student 5's response, and everyone saw student 5's response that said the college should ban alcohol on its campus. Subjects also learned that after the accident, student 5 publicly expressed an opinion that was in line with his private, accidentally revealed response. In particular, subjects learned that student 5 had said that "underage students may feel pressure to drink" if alcohol was allowed on campus, and he added that "drinking is harmful for the development of students' mind and body." In this fashion, the first of the three conditions for

---

[32] In reality, no such discussion format exists. We adapted this format from Study 2 in Willer et al. (2009), in which subjects are in the shoes of one of the students participating in the discussion.

[33] A sample attention-checking question is, "[w]hich of the following best describes the students in the orientation session?" with the correct answer being, "A group of six incoming freshmen participated in the orientation session. They had met each other only briefly before, and they were about to spend the entire year together in the same dorm."

high accusability was introduced: student 5 was revealed publicly as a deviant. At that point, subjects were asked some attention-checking questions.

In the last set of screens, subjects were told the remainder of the story. First, they learned something that did not vary by condition. In particular, they learned that after the discussion, the administrator had announced the results from the initial anonymous questionnaire and noted that two students had initially indicated that the college should ban alcohol. This introduces the second element of a high-accusability situation, in which it is widely known or suspected that there are closeted deviants present. At this point, we added the third element for a high-accusability situation by creating expectations of credible allegations of closeted deviance. We did this by having student 5 say the following in a low voice: "Hmm…if anyone is interested, I'm quite sure I know who initially agreed with me and then wouldn't apparently admit it! That person and I have talked about this stuff before." In addition to the obvious threat of accusation expressed in the comment, we made the threat of accusation more credible by having the known deviant (i.e., student 5) raise the possibility of accusation. Afterward, subjects in the *Baseline* condition were told that the orientation ended here.

*High-Accusability/Mandated and Enforce conditions*. Next, to test our theory, our conditions varied in the degree to which a mandate was provided to assess the deviance. In all *Mandated and Enforce* conditions, subjects were initially told that the "Bloomsbury Ethics Roundtable" format entailed one additional stage in which all students would be called upon in random order to evaluate one another. Accordingly, after student 5 hinted that he knew who the other deviant was (which is when the story ended for subjects in the *Baseline* condition), subjects were told that student 6 was randomly called first to evaluate other students' opinions and that he said the opinions of students 1, 2, 3, and 4 were "very valid," whereas student 5's opinion was

"not valid at all" (see below for more detail on wording). The story ended here for subjects in the *Mandated and Enforce* conditions (i.e., subjects were not told how other students responded to the mandate that was presumably given), and subjects were then asked questions about how they perceived student 6 (see more on the questions below).

   *High-Accusability/Mandated and Not Enforce conditions.* Subjects in the *Mandated and Not Enforce* conditions were also told that students were asked to evaluate one another in random order and that student 6 was called first. In these conditions, subjects were told that student 6 said the opinions of students 1, 2, 3, 4, and 5 are all "equally valid." Just as in the *Mandated and Enforce* conditions, the story ended here for subjects in the *Mandated and Not Enforce* conditions, and as in the *Mandated and Enforce* conditions, subjects were then asked questions about how they perceived student 6.

   *High-Accusability/Entrepreneurial conditions.* By contrast, subjects in the *Entrepreneurial* conditions were not told anything about an additional stage. Instead, after the administrator announced the results from the anonymous questionnaire and student 5 hinted that he knew who the other deviant was (when the story ended for subjects in the *Baseline* condition), subjects were told that "something happened." Subjects were further told that student 6 "suddenly spoke up in a loud voice" and said student 5's opinion was "not valid at all." Afterward, subjects were told the same thing as those in the *Baseline* condition—that the orientation ended and the students were dismissed. The story ended here for the *Entrepreneurial* conditions, and subjects in these conditions were then asked questions about how they perceived all students but student 5 (i.e., students 1 through 4 and 6).

*Different Wordings of the Four Conditions*

The wording used to enforce a norm may influence what motive audiences attribute to an enforcer. Our main intuition was that it may make a difference whether the enforcer simply expresses disapproval for the deviant or whether the enforcer invokes a normative principle. The logic is that it is odd and potentially suspicious for a norm entrepreneur to not justify his or her own enforcement. Therefore, for all sets of conditions other than the *Baseline* condition, we designed three sub-conditions with different wordings. These wordings reflect varying degrees of justification for enforcement and are drafted so that each wording sub-condition has parallel sub-conditions. Overall, subjects were randomly assigned to one of the 10 conditions (3 [*High-Accusability/Mandated and Enforce*, *High-Accusability/Mandated and Not Enforce*, and *High-Accusability/Entrepreneurial*] × 3 [*Simple, Stating Principle*, and *Activist*] + 1 [*High-Accusability/Baseline*]). Table 1 in the online supplement presents enforcement wordings for each condition. We do not state formal hypotheses on how the wording of enforcement matters. Our intuition, however, was that the Simple wordings of enforcement are the least common and plausible, whereas the Activist wordings may be the most plausible and common in norm entrepreneurship.

*Dependent Variable*

Before reading the story, subjects in all conditions had not been told that they would be asked to answer about any specific students. As said before, subjects had only been told that the task is to "assess [their] ability to accurately infer the responses of individuals from a group of college freshmen." After they finished reading the story, subjects in all *Mandated* conditions were asked the following question: "How likely is it that student 6 wrote the following

statements on his anonymous questionnaire?" Subjects were asked to evaluate the statement, "student 6 wrote that 'the college should ban alcohol on its campus'" on scales ranging from 1 (very unlikely) to 10 (very likely). The higher the rating subjects gave to this question the more they doubted student 6's sincerity in his compliance. Because subjects in the *Mandated* conditions were not told of the other students' (i.e., students 1 through 5) evaluations of one another (that presumably happened), subjects were not asked about their guesses on the other students' anonymous questionnaires. That is, subjects in the *Mandated* conditions were explicitly told before answering the questions that they would be asked to only guess student 6's private response and not any other students'. Following the logic explicated below, testing hypothesis 1 involved comparing subjects' responses on student 6 within the *Mandated* conditions (i.e., the *Mandated and Enforce* conditions versus *Mandated and Not Enforce* conditions).

Subjects in the *Entrepreneurial* and *Baseline* conditions were asked to answer the same question but for all five other students (i.e., excluding student 5) because there was no difference in how the administrator had treated those students. Subjects were asked to answer about all five other students in the same screen. Asking about all students in these conditions is especially important because (as explicated below) there are two comparisons that are covered by hypothesis 2: (1) the between-subject comparison (i.e., between student 6 in the *Entrepreneurial* condition and student 6 in the *Baseline* condition) and (2) the within-subject comparison, (i.e., between the enforcer [i.e., student 6] and the bystanders [i.e., students 1, 2, 3, and 4] within the same *Entrepreneurial* conditions). Responses to these questions were used as direct measures of how committed each student appeared because the questions asked subjects about how likely it appeared that the students in question changed their responses. After subjects answered additional questions, including free-response items and demographic items, they were thanked

and given the code to be paid.[34] Finally, because subjects were asked about all five students in both the *Entrepreneurial* and the *Baseline* conditions, we are less concerned that subjects' responses in the *Entrepreneurial* conditions reflect a desire to spot the insincere student any more than subjects' responses in the *Baseline* condition.

*Hypothesis Testing*

Our theoretical framework suggests that subjects draw inferences about others' motives by taking into account a particular feature of social context—that is, whether an explicit mandate for norm enforcement is present or absent. We thus tested hypothesis 1 by comparing the perceived sincerity of enforcers versus non-enforcers within *Mandated* conditions; and we tested hypothesis 2 by comparing the perceived sincerity of enforcers versus non-enforcers within *Not Mandated* conditions. The results of these tests are, therefore, down each column in Table 2 and never diagonal or within each row. Note that the very contextual feature we vary also changes how enforcers may be compared with non-enforcers. In particular, when there is no mandate, all individuals have exactly the same opportunity to engage in norm enforcement; as such, any entrepreneurial enforcer's sincerity can be meaningfully compared to all others who had the same opportunity to interrupt the flow of social interaction and enforce but did not do so. Therefore, for hypothesis 2, we expect that subjects will perceive the enforcer in the *Entrepreneurial* conditions (i.e., student 6 in the *Entrepreneurial* conditions) to be less sincere than bystanders in any condition that lacks a mandate: either the bystanders in the

---

[34] No demographic factors except for gender affected subjects' doubts about student 6's commitment: men were in general more suspicious than women that student 6 changed his opinion. However, this effect did not substantively change the main effects of our manipulations, so we do not discuss this further.

*Entrepreneurial* conditions (i.e., students 1 through 4 in the *Entrepreneurial* conditions) or the bystanders in the *Baseline* condition (e.g., student 6 in the *Baseline* condition).

By contrast, insofar as mandates are separately given to all would-be enforcers, their opportunities to enforce are not equivalent. In particular, any responses to a mandate that presumably followed student 6's response to the mandate (i.e., responses to the mandate by the students 1 through 5) might be interpreted as influenced by student 6's response to the mandate. As such, the appropriate way to compare enforcers and non-enforcers in the *Mandated* conditions is through a between-subject test, whereby the sincerity of the first person who was given a mandate and enforced (i.e., student 6 in the *Mandated and Enforce* conditions) is compared to the sincerity of another person who was given the mandate first in the group and did not enforce in response to that mandate (i.e., student 6 in the *Mandated and Not Enforce* conditions). Tests for hypothesis 1, therefore, involve only between-subject comparisons.

**Study 1 Results**

*Validating Hypothesis 1*

We argued that the illusion of sincerity would be generated when enforcement is in response to a mandate because non-enforcers would lack plausible deniability that they had witnessed or recognized the significance of the deviance. This is reflected in hypothesis 1. In order to validate this prediction, we developed a design that should replicate results from Willer et al. (2009), which showed a higher perception of sincerity for the enforcer than for someone was also given a mandate but did not enforce. Our results successfully replicated the findings of Willer et al. (2009). Figure 1 shows that student 6 in the *High-Accusability/Mandated and Enforce* conditions appeared less likely to have changed his response (M = 4.62) than did student

6 in the *High-Accusability/Mandated and Not Enforce* condition (M = 6.74; t = $5.87^{35}$; $p < 0.001$; degrees of freedom [df] = 307). [36] In addition, an analysis of variance (ANOVA) indicates that suspicion was elicited mainly by the refusal to enforce and not by a particular wording sub-condition (F[1,301] = 35.33; $p < 0.001$; see Table 2 in the online supplement).

*Validating Hypothesis 2*

Having replicated the finding (from Willer et al. 2009) that the illusion of sincerity can be elicited by those who enforce in response to a mandate, we then tested whether audience suspicion is aroused when enforcers lack such a mandate. As the main test of our argument, hypothesis 2 predicted that because an entrepreneurial enforcer could have ignored the deviance and enjoyed plausible deniability, he would face greater suspicion of deviance than would bystanders. The results validate this prediction. Student 6 in the *High-Accusability/Entrepreneurial* conditions was perceived as more likely to have changed his response than students 1, 2, 3, and 4 on average (M = 6.02 vs. M = 3.83; t = 5.86; $p < 0.001$; df = 298). Results are substantively the same in each *High-Accusability/Entrepreneurial* wording sub-condition. By contrast, there is no evidence that student 6 in the *High-Accusability/Baseline* condition seemed any more suspicious than students 1, 2, 3, and 4 on average (M = 4.77 vs. M = 4.16; t = 1.28; $p = 0.20$; df = 98), indicating that it is entrepreneurial enforcement (and not other

---

[35] Although the t-test is efficient and easy to interpret, it also assumes equal variance between two populations of comparison. For a robustness check, we also use the Mann–Whitney test that does not assume equal variance, and all results remain substantively the same (Fay and Proschan 2010; Mann and Whitney 1947).

[36] All statistical tests in this article were done in *R*. All ANOVA tests in this article are type II unbalanced ANOVA because each wording sub-condition does not always have an equal number of subjects. Unbalanced ANOVA was run by using the "car" package.

characteristics of student 6) that made him appear more suspicious. Figure 2 shows these results visually.

The between-condition comparison (i.e., student 6 in the *High-Accusability/Entrepreneurial* conditions vs. student 6 in the *High-Accusability/Baseline* condition) further demonstrates that entrepreneurial enforcement invites suspicions of deviance. This comparison is a particularly conservative test because the experiment was set up so that student 6 in the *High-Accusability/Baseline* condition (as well as in other conditions) went last in the public discussion. Insofar as subjects were told there was a closeted deviant and asked to report their suspicions of deviance, student 6 in the *High-Accusability/Baseline* condition might suffer from higher suspicion of deviance to begin with, especially because subjects are implicitly trying to pinpoint one closeted deviant. Nevertheless, when student 6 remained silent like those before him, he elicited significantly less suspicion than in the condition when he engaged in entrepreneurial enforcement (M for student 6 in all *High-Accusability/Entrepreneurial* conditions = 6.02 vs. M for student 6 in the *High-Accusability/Baseline* condition = 4.77; $t = 2.51$; $p = 0.01$; df = 198). The ANOVA results further confirm that entrepreneurially enforcing elicits suspicion ($F[1,183] = 4.23$, $p < 0.05$; see Table 3 in the online supplement) and that different wording sub-conditions matter little. Again, Figure 1 visually shows these results.

## Study 1 Discussion

In study 1, we validated hypothesis 1 by replicating the illusion of sincerity effect (Willer et al. 2009) showing that when enforcement occurred in response to a mandate; in such situations, the ulterior motive inference is less salient even in a situation of high accusability. But supporting hypothesis 2, study 1 also shows that suspicions are significantly aroused when

enforcement occurs via entrepreneurial initiative. In the absence of a mandate, plausible

deniability that one did not see or appreciate the deviance is more available to a bystander; as

such, entrepreneurial enforcement is suspicious.

It is worth clarifying that although results from both the within-condition comparison and

between-condition comparison validate hypothesis 2, the more useful comparison is likely the

within-condition comparison. We limited the comparison of the *High-Accusability/Mandated*

conditions to be between student 6's of the two different conditions in order to control for factors

that are less applicable to the real world (i.e., the unlikely random order in which one is

mandated to enforce).[37] Yet counterfactual scenarios, by definition, do not exist in the real world.

Instead, one might more often be rewarded or punished on the basis of one's appearance of

commitment in comparison to others' in the same situation. Based on such reasoning, we regard

the within-condition effect shown in Figure 2 from the *Entrepreneurial* conditions as the

strongest evidence in support of hypothesis 2. Consequently, the problem with entrepreneurial

enforcement under high accusability is that one invites suspicion of oneself in comparison with

others in the same situation.

However, study 1 alone cannot test whether high accusability is a necessary situational

feature for enforcement to elicit audience suspicion. Insofar as our theory hinges on the

assumption that audiences perceive ulterior motives of covering deviance from entrepreneurial

enforcement, we need to directly test that suspicion would be attenuated when there is little

reason to defend oneself. Such a test is especially needed because one might argue that the

audience suspicion, elicited by entrepreneurial enforcement in study 1, can be aroused by any

interruptions and not necessarily by acts suspected as deriving from ulterior motives. That is,

---

[37] We further discuss this point in the Methodological Appendix in the online supplement.

insofar as subjects could have a desire to spot the student most likely to be insincere, any attention generated by the interruption (including entrepreneurial enforcement) might have elicited suspicion. Study 2, therefore, is designed to test the idea that even when subjects might have the same level of desire to spot the student most likely to be insincere, entrepreneurial enforcement does not elicit suspicion in the absence of accusability. Therefore, we now move to study 2, in which we introduced conditions that parallel those of study 1 but in low-accusability situations. This allowed us to test hypothesis 3, which holds that entrepreneurial enforcement is more likely to create the impression that the enforcer endorses the norm in low-accusability situations (as implemented in study 2) than in high-accusability situations (conditions from study 1).

## Study 2 Design and Procedures

### Recruitment and Conditions

Five hundred seventy-four subjects were recruited via Amazon Mechanical Turk in the same time period as study 1 and paid the same as those in study 1. Study 1 and study 2 were conducted within days, and participants in study 1 were excluded, making them two comparable samples—essentially two sets of sub-conditions of the same experiment. Because the primary purpose of adding study 2 was to reduce accusability and, therefore, the strategic value behind enforcement, conditions in study 2 are parallel to study 1 but only in low-accusability situations. Therefore, the four sets of conditions of study 2 are as follows: *Low-Accusability/Mandated and Enforce, Low-Accusability/Mandated and Not Enforce, Low-Accusability/Entrepreneurial*, and *Low-Accusability/Baseline*. Again, subjects were assigned to only one of these conditions, and they were not aware of the presence of other conditions; the full set of conditions is again

summarized in Table 2. Enforcement wordings for each condition remain exactly the same as in study 1.

*Study 2 Vignette*

Because the only difference between the study 1 and study 2 conditions is the level of accusability, the conditions of study 2 are different from their corresponding conditions of study 1 in only two subtle ways. First, while subjects in study 1 were told that students "had met each other only briefly before," subjects in study 2 were told that students "had never met" one another before the orientation. This change subtly informs subjects in study 2 that students could not make educated guesses about one another's true commitment based on their experiences, thereby lowering the threat of credible accusation. Second, subjects no longer heard one of the students (i.e., student 5) claim that he knew who privately supported banning alcohol (i.e., who does not endorse the norm). To recall, subjects in all conditions of study 1 were told the following: after the public discussion in which all students but student 5 expressed their opposition to banning alcohol, the students were informed that there was another student who had initially favored banning alcohol on campus; then, student 5 said he was quite sure who was the closeted deviant, thereby raising the threat of accusation. By contrast, subjects in study 2 were not told of any comment after the public announcement and instead assumed that the students proceeded right to the next stage. In other words, right after the public announcement about a closeted deviant by the administrator, subjects in the *Low-Accusability/Mandated* conditions were told that student 6 was called in random order to evaluate the other students. After student 6 evaluated the other students, subjects were asked the same questions as ones subjects in the *High-Accusability/Mandated* conditions of study 1 were asked. In the *Low-*

*Accusability/Entrepreneurial* conditions, subjects were told that student 6 suddenly spoke up to condemn student 5 (i.e., the known deviant), and subjects were also told that the orientation ended there. Subjects in the *Low-Accusability/Baseline* condition were told that the orientation ended right after the administrator made the announcement. Subjects in each condition of study 2 were asked the exact same questions as subjects in the corresponding condition of study 1.

**Study 2 Results**

*Validating Hypothesis 3*

The main objective of study 2 was to test whether an entrepreneurial enforcer in a situation of low accusability (i.e., student 6 in *Low-Accusability/Entrepreneurial* conditions in study 2) would appear less suspicious than the equivalent one in a situation of high accusability (i.e., student 6 in *High-Accusability/Entrepreneurial* conditions in study 1). Our results support this prediction. That is, suspicion is considerably greater for the entrepreneurial enforcer in high-accusability situations as compared to the same enforcer in low-accusability situations (M for student 6 in all study 1 *High-Accusability/Entrepreneurial* conditions = 6.02 vs. M for student 6 in all study 2 *Low-Accusability/Entrepreneurial* conditions = 4.49; t = 3.66; $p < 0.001$; df = 316). We also used ANOVA to identify the effect of the main manipulation, and results confirm the effect of accusability (F[1,302] = 13.49; $p < 0.001$; see Table 4 in the online supplement). In addition, in the *Low-Accusability/Entrepreneurial* conditions, there is no evidence that audiences suspected any problematic commitment from student 6, who took entrepreneurial initiative to enforce the norm (M = 4.49) compared to students 1, 2, 3, and 4 on average (M = 4.26; t = 0.72; $p = 0.48$; df = 334). It is also important to note that the additional suspicion directed toward student 6 (compared to that on students 1, 2, 3, and 4) in the *Low-*

*Accusability/Baseline* condition (M for student 6 = 5.47 vs. M for students 1, 2, 3, and 4 = 4.73; t

= 1.63; $p$ = 0.11; df = 114) was as large as or larger than the additional suspicion directed toward

student 6 in the *Low-Accusability/Entrepreneurial* conditions, indicating that entrepreneurial

enforcement in low-accusability situations did little to elicit additional suspicion. Figure 3 shows

these results visually.

The between-condition comparisons (i.e., student 6 in the *Low-*

*Accusability/Entrepreneurial* conditions vs. student 6 in the *Low-Accusability/Baseline*

condition) further validate that entrepreneurial enforcement did not elicit further doubts about the

enforcer's commitments in a situation of low accusability. Figure 4 shows this comparison

visually. In fact, if student 6 enforced entrepreneurially and did so in what appeared to be a

plausible manner, he aroused significantly less suspicion than student 6 in the *Baseline*

condition. In particular, student 6 in the *Low-Accusability/Entrepreneurial–Stating Principle*

sub-condition appeared significantly less suspicious (M = 4.16) than student 6 in the *Baseline*

condition (M = 5.47; t = 2.20; $p$ < 0.05; df = 226). Student 6 in the *Low-*

*Accusability/Entrepreneurial–Activist* sub-condition also appeared significantly less suspicious

(M = 3.84 vs. M = 5.47; t = 2.73; $p$ < 0.01; df = 228). The ANOVA test further confirmed these

results (F[1,222] = 6.61; $p$ = 0.01; see Table 5 in the online supplement[38]).

*Replicating Willer et al. (2009) in a Situation of Low Accusability*

Lastly, it is useful to confirm that enforcement in response to a mandate in a situation of

low accusability can still fend off potential suspicion that the entrepreneurial enforcement in a

---

[38] The significant interaction effect from Table 5 in the online supplement suggests that the main effect of entrepreneurial enforcement is significantly weaker for the *Entrepreneurial–Simple* sub-condition. This is in line with our intuition that entrepreneurially enforcing in an implausible manner not invoking a normative principle does not enhance one's perception of commitment.

situation of high accusability elicits, as depicted in Table 1. Study 2 does this. Subjects in general thought that student 6 in the *Low-Accusability/Mandated and Enforce* conditions was less likely to have changed his response (M = 4.23) than student 6 in the *Low-Accusability/Mandated and Not Enforce* conditions (M = 6.32; t = 6.26; $p < 0.001$; df = 346; see Figure 4). The ANOVA that compared the two mandated conditions in a situations of low accusability further validated that the effect is significant at the conventional level (F[1, 342] = 36.33; $p < 0.001$; see Table 6 in the online supplement).

## Study 2 Discussion

Study 2 tests hypothesis 3 and demonstrates that suspicions that are elicited by entrepreneurial enforcement in the general situation of high accusability can be "turned off" by lowering the level of accusability. Qualitative responses support this part of our theory as well. After asking questions for our main dependent variable and the suspicion score, we asked subjects "why [they] evaluated student 6 the way [they] did." Their answers suggest that they recognize entrepreneurial enforcement has no particular strategic value in a low-accusability situation: as one subject put it, "[student 6] seemed very passionate about his opinion. There was no need to speak out against student 5 unless it really bothered him."

Another important implication of study 2 is that simply interrupting the flow of social interaction does not raise the suspicion of deviance. In fact, the between-condition results showed that when one interrupted the flow of social interaction and enforced in a plausible manner in a situation of low accusability (i.e., in the *Low-Accusability/Entrepreneurial–Stating Principle* and the *Low-Accusability/Entrepreneurial–Activist* conditions), one appeared less suspicious than someone who remained silent. These results then demonstrate that the suspicion

elicited by entrepreneurial enforcement in *High-Accusability* conditions in study 1 was not simply an experimental byproduct of study 1 subjects' desires to spot the insincere student—both high accusability and entrepreneurial enforcement are needed to elicit suspicion.

Yet although these results from a situation of low accusability show that enforcement through interruption alone is not sufficient to raise suspicion, they do not necessarily mean that all interruptions are the same. In fact, we designed a posttest in which student 5 was in a situation of high accusability and said, "You know, I really like this Bloomsbury Roundtable—this is exactly the kind of experience I wanted to have in college. I'm so glad that I got to be part of it, and I'm looking forward to the rest of the orientation." The objective of this test was to see whether mere interruption using an orthogonal statement might raise suspicion as much as the interruption via enforcement. Results show that entrepreneurial norm enforcement indeed elicited less suspicion than did this irrelevant comment. This suggests that although entrepreneurial enforcement in a situation of high accusability does make the ulterior motive inference more salient, it still keeps the sincere motive inference viable by at least claiming sincerity, whereas someone who interrupts the flow of social interaction without any defensive acts (e.g., enforcement) might not keep afloat even that much of the sincere motive inference. Results from study 2 and the posttest then highlight the situational features that enable enforcement to generate the illusion of sincerity.

## Limitations

Before we discuss and draw implications from our theory and results, it is worth clarifying an important limitation of our study design. Our vignette experiments entailed informing subjects about a hypothetical scenario in which normative deviance and enforcement

occurred; we then asked subjects to judge the enforcer's (and bystanders') sincerity. Although results from these vignette studies demonstrated when the enforcer appears suspicious of having an ulterior motive, our subjects were not themselves in the position to enforce (or remain silent). Such a setup is difficult to implement in online experiments in which subjects are anonymous – i.e., when subjects do not have reasons to care about how sincere their conformity appears. Without such a design, however, we do not yet know if would-be enforcers act in the manner implied by the pattern of audience response about which we have theorized and provided experimental results. Therefore, we focused on addressing why the perception of suspicion is elicited by enforcement (and not necessarily why norms are under-enforced), although a possible implication of our theory is that norms will be under-enforced.

At the same time, the experimental paradigm we developed does provide important indirect evidence as to why norms might be under-enforced when entrepreneurial initiative is required to enforce that norm. As Willer et al. (2009:477) put it, "[i]f 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." Sociologists have indeed long documented how perceptions by others drive individuals' behaviors (e.g., Bourdieu 1984; Goffman 1959), and our theory and empirical results similarly identify conditions under which suspicions aroused by entrepreneurial norm enforcement might lead to the under-enforcement of norms. We discuss this issue further in the General Discussion section below.

It is also worth emphasizing the weaknesses and strengths of our sampling strategy. Given that our studies feature the issue of drinking on college campuses, undergraduate participants might be more knowledgeable about the vignette setting. However, we were also

concerned that their close involvement in campus life might influence their responses and constrain us from controlling our treatments. Insofar as our sample is familiar (either through experience or social learning) with the setting presented in our vignette and is more representative of the general population than college students in lab settings (Berinsky et al. 2012; Buhrmester et al. 2011), conducting vignette experiments on this online platform seems particularly well suited to test our theory while maximizing our control over our treatments of interest (Parigi, Santana, and Cook 2017).

Finally, a limitation of our sampling strategy is that the samples were limited to the United States–based subjects. We do not have theoretical reasons for entrepreneurial enforcement to arouse suspicion only among the U.S.-based subjects, yet we are limited by substantive (e.g., knowledge around a setting in which such contestation of a norm is common) and practical (e.g., tools with which to collect samples that are as general as those of U.S. equivalents on the Amazon Mechanical Turk platform) constraints. Empirical accounts documented from political purges (e.g., the Chinese Cultural Revolution) and lay beliefs from different cultures (e.g., sayings analogous to "Whoever smelt it dealt it") suggest that this theoretical mechanism is prevalent even though we do not validate this theory with samples other than the U.S.-based ones in this article.

## GENERAL DISCUSSION

Recent research has suggested how one might use "false enforcement" (Willer et al. 2009) to mask hidden deviance and that this is an important basis for upholding norms. However, various empirical accounts and lay beliefs suggest that audiences are often suspicious of the ulterior motives that this strategy implies. The question that emerges from this tension then

is why and when enforcement can create an illusion of sincerity and why and when it elicits suspicions of ulterior motives.

The main contribution of this article is to identify two situational features that explain variation in suspicion that a norm enforcer is a hidden deviant: (1) whether actors face a significant threat of facing credible accusations of deviance and thus pressure to feign commitment and (2) whether the context is such that actors have no mandate to engage in enforcement but must take entrepreneurial initiative to do so. Our experiments replicated past results showing that deviance is successfully masked when actors are given a mandate to engage in enforcement. But our results also support our argument that suspicion is highest in the more common case when actors have no mandate and precisely when false enforcement is attractive as a strategy to mask hidden deviance—i.e., under high accusability. Again, the general logic is that someone who takes entrepreneurial initiative to sanction deviance could have ignored the deviance without having endorsed it; this invites the suspicion that she has a special motive to signal her commitment to the norm. By contrast, someone who fails to condemn a norm violation when given a mandate to do so effectively endorses the deviance and thus contributes to the undermining of the norm.

We now put the contribution of our article in a broader context by considering our theory's implications for how norm enforcement might evolve over time in light of situations documented in previous studies. We then conclude by highlighting the tension between rewards and punishments faced by actors who undertake prosocial action.

## Accounting for Persistence of Norm Entrepreneurship and Its Implications

Our theory and empirical tests pointed to conditions under which norm enforcers can escape suspicions of ulterior motive even in a situation of high accusability—i.e., when assessment is mandated—as documented by Willer et al. (2009). Although our conjecture is that such instances are rare, it is nevertheless important to note empirical cases that meet this condition reaffirming the literature's and our evidence's external validity. In particular, Adut's (2004) analysis of the French investigating magistrates' anticorruption campaign is a case of enforcement in response to a mandate. Because the job of the investigating magistrates was to legally investigate anyone who was suspected of illegality (corruption, more specifically, in Adut's case), they could escape suspicions of ulterior motives for their enforcement, similar to the enforcer in our *Mandated and Enforce* conditions. Accordingly, Adut (2004:547) observed that the French investigating magistrates in the 1990s were able to seize their "opportunities for public displays of moral rectitude and courage" and enhance their status by accusing then–high-status political actors of being corrupt. Their institutional role gave them a mandate, which helped the sincere motive inference become more viable (also see Erikson 1966; Evans-Pritchard 1976; Reilly 2016).[39] Yet the case of French magistrates seems to be a rather special case. It is difficult to come up with other examples in which enforcers are so well protected from

---

[39] Erikson's analysis of the Salem witch trials (1966; also, Evans-Pritchard [1976]) also documents how adolescent girls who were already identified as "witch doctors" were regarded as reliable when they identified other girls as witches. The designated role provided them with a mandate to assess others' professed sincerity in upholding the norm (of not being a witch), and they could escape suspicions of ulterior motives for hiding their own deviance and protect their status in the community. See also Reilly (2016:25–8) for an intriguing contrast between entrepreneurial enforcement and enforcement in response to a mandate in a community of stand-up comedians. He describes how Joe Rogan—who entrepreneurially accused Carlos Mencia of joke theft—was often criticized for his "opportunistic and self-promotional grandstanding," whereas Marc Maron—who asked Mencia to address the issue on his interview podcast—was able to escape such criticism because he had little to gain and the topic was "seemingly unavoidable."

suspicions of ulterior motive by a mandate, as even enforcers occupying such institutional roles are rarely given a mandate to respond to specific norm violations. It seems much more common for enforcement to require some degree of entrepreneurial initiative, as is highlighted by Becker's (1963) classic insight, and as shown in our results, suspicions may be aroused if an enforcer is viewed as straying beyond the mandate.

It is also worth noting that enforcement is most attractive as a strategy for signaling sincere conformity in situations of high accusability, and therefore, it is most pertinent to investigate what inference audiences draw in such situations. Insofar as the individual motivation for enforcement is about signaling sincere conformity, enforcement has little value in situations of low accusability: why incur the personal cost and enforce when there is little reason to portray one's conformity as sincere (see also Centola et al. 2005:1031)?[40] Consequently, investigating the perception of enforcers is most appropriate in situations of high accusability where enforcement is most attractive as a strategy, and our results demonstrate that the most common form of norm enforcement—i.e., entrepreneurial enforcement—elicits the suspicion of deviance in such situations. The larger implication of our theory is that insofar as the enforcement of a norm requires entrepreneurial initiative, the norm may be under-enforced—despite its potential value as a commitment-enhancing strategy—and ultimately fail to govern social action.

Yet we recognize that enforcement does not necessarily cease under such conditions. The empirical examples, such as the Chinese Cultural Revolution, attest to suspicion for enforcement only because enforcement occurred despite significant suspicion. This suggests that there (at least sometimes) occurs the "breakout of enforcement" in scandals (Adut 2008) and moral panic

---

[40] Obviously, we are not barring exceptional cases in which individuals might be personally invested in the issue to incur a personal cost and enforce the norm regardless of public perception, although those individuals might still refrain from entrepreneurially enforcing lest they be regarded as opportunists.

(Goode and Ben-Yehuda 1994).[41] Then, when the context is one of high accusability and requires entrepreneurial initiative to enforce, why might enforcement still continue? Addressing this question is a fruitful avenue for future research: although this article's implication is that norms will be under-enforced among strategic actors behaving rationally; with full information on the costs of enforcement, such restrictive assumptions of the intent, rationality, and knowledge of the actors might not apply in other situations. An examination of why and when enforcement breaks out in high-accusability situations, therefore, should encompass a diverse array of strategic and behavioral tendencies.[42]

## CONCLUSION

Previous research has suggested that contributions to a public good can be driven by selective reputational incentives, and contributions, such as prosocial behaviors, are more likely to occur because of them (e.g., Willer 2009). However, more recent research suggests that an overt pursuit of the reputational benefits may lower the perception of commitment (e.g., Hahl and Zuckerman 2014; Simpson and Willer 2008), and this article identified conditions under which a contributor to a public good (i.e., an enforcer) can plausibly deny suspicions of an ulterior motive. This article thus suggests that situational features may be able to heighten the perception that prosocially oriented actors are seeking reputational or material benefits rather than the public good, and this in turn might make prosocial behavior less likely even in the presence of selective incentives for prosocial behavior. Future research should investigate (1) how such suspicions of ulterior motives affect different prosocial behaviors and (2) why some prosocial behaviors persist

---

[41] Another example comes from the scandal surrounding Russia's meddling in the 2016 U.S. presidential election, in which Donald J. Trump seems to be using an accusation of the deviance in order to defend himself (Phillips 2017).
[42] We thank the editor for this suggestion.

despite suspicions of ulterior motives. By addressing these questions, we will be able to more fully account for conditions under which prosocial behaviors persist and conditions under which they do not.

## REFERENCES

Adut, Ari. 2004. "Scandal as Norm Entrepreneurship Strategy: Corruption and the French Investigating Magistrates." *Theory and Society* 33 (5): 529–78.

———. 2005. "A Theory of Scandals: Victorians, Homosexuality, and the Fall of Oscar Wilde." *American Journal of Sociology* 111(1): 213-248.

———. 2008. *On Scandal: Moral Disturbances in Society, Politics, and Art.* Cambridge: Cambridge University Press.

———. 2012. "A Theory of the Public Sphere." *Sociological Theory* 30 (4): 238–62.

Alchian, Armen A., and Harold Demsetz. 1972. "Production, Information Costs, and Economic Organization." *The American Economic Review* 62 (5): 777–95.

Asch, Solomon E. 1951. Effects of Group Pressure on the Modification and Distortion of Judgments. In *Groups, Leadership and Men*, edited by H. Guetzkow, 177–190. Pittsburgh, PA: Carnegie Press.

Becker, Howard S. 1960. "Notes on the Concept of Commitment." *American Journal of Sociology* 66 (1): 32–40.

———. 1963. *Outsiders: Studies in The Sociology Of Deviance.* New York: Free Press.

Berinsky, Adam J., Gregory A. Huber, and Gabriel S. Lenz. 2012. "Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk." *Political Analysis* 20 (3): 351–68.

Bourdieu, Pierre. 1984. *Distinction: A Social Critique of the Judgement of Taste.* Harvard University Press.

Brañas-Garza, Pablo, Antonio M. Espín, Filippos Exadaktylos, and Benedikt Herrmann. 2014. "Fair and Unfair Punishers Coexist in the Ultimatum Game." *Scientific Reports* 4 (August).

Buhrmester, Michael, Tracy Kwang, and Samuel D. Gosling. 2011. "Amazon's Mechanical Turk A New Source of Inexpensive, Yet High-Quality, Data?" *Perspectives on Psychological Science* 6 (1): 3–5.

Carlos, W. Chad, and Ben W. Lewis. 2017. "Strategic Silence: Withholding Certification Status as a Hypocrisy Avoidance Tactic." *Administrative Science Quarterly* Online First: 1-40.

Centola, Damon, Robb Willer, and Michael Macy. 2005. "The Emperor's Dilemma: A Computational Model of Self-Enforcing Norms." *American Journal of Sociology* 110 (4): 1009–40.

Cheng, Nien. 1986. *Life and Death in Shanghai.* London: Grafton Books.

Coleman, James. 1990. *Foundations of Social Theory.* Belnap Press.

Di Stefano, Giada, Andrew A. King, and Gianmario Verona. 2015. "Sanctioning in the Wild: Rational Calculus and Retributive Instincts in Gourmet Cuisine." *Academy of Management Journal* 58 (3): 906-931.

Duncan, Jericka. 2015, "Will Dartmouth's Hard Alcohol Ban Make Students Safer?" Accessed January 6. http://www.cbsnews.com/news/will-dartmouths-hard-alcohol-ban-make-students-safer/.

Evans-Pritchard, Edward E. 1976. *Witchcraft, Oracles and Magic among the Azande.* Oxford: Oxford University Press.

Erikson, Kai. 1966. *Wayward Puritans: A Study in the Sociology of Deviance.* Boston: Prentice Hall.

Fay, Michael P., and Michael A. Proschan. 2010. "Wilcoxon-Mann-Whitney or T-Test? On Assumptions for Hypothesis Tests and Multiple Interpretations of Decision Rules." *Statistics Surveys* 4: 1–39.

FeldmanHall, Oriel, Peter Sokol-Hessner, Jay J. Van Bavel, and Elizabeth A. Phelps. 2014. "Fairness Violations Elicit Greater Punishment on Behalf of Another than for Oneself." *Nature Communications* 5 (October): 5306.

Fehr, Ernst, and Simon Gächter. 2002. "Altruistic Punishment in Humans." *Nature* 415 (6868): 137–40.

Fine, Gary Alan. 1996. "Reputational Entrepreneurs and the Memory of Incompetence: Melting Supporters, Partisan Warriors, and Images of President Harding." *American Journal of Sociology* 101.5 (March): 1159-1193.

Gilbert, Daniel T., and Patrick S. Malone. 1995. "The Correspondence Bias." *Psychological Bulletin* 117 (1): 21–38.

Goffman, Erving. 1959. *The Presentation of Self in Everyday Life*. Anchor Books.

Goode, Erich and Nachman Ben-Yehuda. 1994. "Moral Panics: Culture, Politics, and Social Construction." *Annual Review of Sociology* 20: 149-71.

Hahl, Oliver, and Ezra W. Zuckerman. 2014. "The Denigration of Heroes? How the Status Attainment Process Shapes Attributions of Considerateness and Authenticity." *American Journal of Sociology* 120 (2): 504–54.

Herrmann, Benedikt, Christian Thöni, and Simon Gächter. 2008. "Antisocial Punishment Across Societies." *Science* 319 (5868): 1362–67.

Head, John G. 1974. *Public Goods and Public Welfare*. Durham, NC: Duke University Press.

Henrich, Joseph, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr, Herbert Gintis, and Richard McElreath. 2001. "In Search of Homo Economicus: Behavioral Experiments in 15 Small-Scale Societies." *The American Economic Review* 91 (2): 73–78.

Homans, George C. 1950. *The Human Group*. Brace & Company, New York.

Horne, Christine. 2001. "The Enforcement of Norms: Group Cohesion and Meta-Norms." *Social Psychology Quarterly* 64 (3): 253–66.

———. 2007. "Explaining Norm Enforcement." *Rationality and Society* 19 (2): 139–70.

Humphreys, Laud. 1970. *Tearoom Trade: Impersonal Sex in Public Places*. Duckworth.

Ingram, Paul, Lori Qingyuan Yue, and Hayagreeva Rao. 2010. "Trouble in Store: Probes, Protests, and Store Openings by Wal-Mart, 1998–2007." *American Journal of Sociology* 116 (1): 53–92.

Jones, Edward E, and Victor A Harris. 1967. "The Attribution of Attitudes." *Journal of Experimental Social Psychology* 3 (1): 1–24.

Jordan, Jillian J., Moshe Hoffman, Paul Bloom, David G. Rand. 2016. "Third-Party Punishment as a Costly Signal of Trustworthiness." *Nature*.

Kim, Minjae. 2018. "A Man is Known by His Cup: Signaling Commitment via Insincere Conformity." Working Paper.

King, Brayden G., and Nicholas A. Pearce. 2010. "The Contentiousness of Markets: Politics, Social Movements, and Institutional Change in Markets." *Annual Review of Sociology* 36 (1): 249–67.

Kuran, Timur. 1995. *Private Truths, Public Lies: The Social Consequences of Preference Falsification*. Cambridge, Mass.: Harvard University Press.

Mann, Henry B., and Douglas R. Whitney. 1947. "On a Test of Whether One of Two Random Variables Is Stochastically Larger than the Other." *The Annals of Mathematical Statistics* 18 (1): 50–60.

Mason, Winter, and Siddharth Suri. 2011. "Conducting Behavioral Research on Amazon's Mechanical Turk." *Behavior Research Methods* 44 (1): 1–23.

McDonnell, Mary-Hunter. 2015. "Radical Repertoires: The Incidence and Impact of Corporate-Sponsored Social Activism." *Organization Science* 27 (1): 53–71.

McDonnell, Mary-Hunter, and Brayden King. 2013. "Keeping up Appearances Reputational Threat and Impression Management after Social Movement Boycotts." *Administrative Science Quarterly* 58 (3): 387–419.

Miller, Dale T., and Cathy McFarland. 1991. "When Social Comparison Goes Awry: The Case of Pluralistic Ignorance." Pp. 287–313 in *Social Comparison: Contemporary Theory and Research*, edited by J. Suls and T. A. Wills. Hillsdale, N.J.: Erlbaum.

Oliver, Pamela. 1980. "Rewards and Punishments as Selective Incentives for Collective Action: Theoretical Investigations." *American Journal of Sociology* 85 (6): 1356–75.

Olson, Mancur. 1971. *The Logic of Collective Action: Public Goods and the Theory of Groups.* Cambridge, Mass.: Harvard University Press.

Parigi, Paolo, Jessica J. Santana, and Karen S. Cook. 2017. "Online Field Experiments: Studying Social Interactions in Context." *Social Psychology Quarterly* 80 (1): 1–19.

Parsons, Talcott. 1937. *The Structure of Social Action.* The Free Press.

Phillips, Amber J. 2017. "Trump's Defense against Comey Has Fallen into a Predictable Pattern: Make a Baseless Accusation." *Washington Post.* Accessed June 12. https://www.washingtonpost.com/news/the-fix/wp/2017/06/12/trumps-defense-against-comey-has-fallen-into-a-predictable-pattern-make-a-baseless-accusation/.

Phillips, Damon J., Catherine J. Turco, and Ezra W. Zuckerman. 2013. "Betrayal as Market Barrier: Identity-Based Limits to Diversification among High-Status Corporate Law Firms." *American Journal of Sociology* 118 (4): 1023–54.

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.

Rao, Hayagreeva, Lori Qingyuan Yue, and Paul Ingram. 2010. "Laws of Attraction: Regulatory Arbitrage in the Face of Activism in Right-to-Work States." *American Sociological Review* 76.3 (2011): 365-385.

Reilly, Pat. 2016. "No Laughter Among Thieves: Informal Intellectual Property Rights and Status Dynamics in Stand-Up Comedy." Working Paper

Raihani, Nichola J., and Redouan Bshary. 2015. "The Reputation of Punishers." *Trends in Ecology & Evolution* 30 (2): 98–103. Rowley, Charles Kershaw; Peacock, Alan T. 1975. *Welfare Economics: A Liberal Restatement.* First Edition. London: Martin Robertson.

Samuelson, Paul A. 1954. "The Pure Theory of Public Expenditure." *The Review of Economics and Statistics* 36 (4): 387–89.

Schelling, Thomas. *The Strategy of Conflict.* 1980. Cambridge: Harvard University Press.

Simpson, Brent, Ashley Harrell, and Robb Willer. 2013. "Hidden Paths from Morality to Cooperation: Moral Judgments Promote Trust and Trustworthiness." *Social Forces* 91 (4): 1529–48.

Simpson, Brent, and Robb Willer. 2008. "Altruism and Indirect Reciprocity: The Interaction of Person and Situation in Prosocial Behavior." *Social Psychology Quarterly* 71 (1): 37–52.

————. 2015. "Beyond Altruism: Sociological Foundations of Cooperation and Prosocial Behavior." *Annual Review of Sociology* 41 (1): 43–63.

Sunstein, Cass R. 1996. "Social Norms and Social Roles." *Columbia Law Review* 96 (4): 903–68.

Van Maanen, John. 1973. "Observations on the Making of Policemen." *Human Organization* 32 (4): 407–18.

Walder, Andrew G. 2006. "Ambiguity and Choice in Political Movements: The Origins of Beijing Red Guard Factionalism." *American Journal of Sociology* 112 (3): 710–50.

————. 2009. *Fractured Rebellion: The Beijing Red Guard Movement*. Cambridge, Mass.; London: Harvard University Press.

Wetts, Rachel, and Robb Willer. 2015. "Anxiety and Austerity: A Group Position Account of White Americans' Opposition to Welfare." Working Paper.

Willer, Robb. 2009. "Groups Reward Individual Sacrifice: The Status Solution to the Collective Action Problem." *American Sociological Review* 74 (1): 23–43.

Willer, Robb, Ko Kuwabara, and Michael W. Macy. 2009. "The False Enforcement of Unpopular Norms." *American Journal of Sociology* 115 (2): 451–90.

Willer, Robb, Christabel L. Rogalin, Bridget Conlon, and Michael T. Wojnowicz. 2013. "Overdoing Gender: A Test of the Masculine Overcompensation Thesis." *American Journal of Sociology* 118 (4): 980–1022.

Yoon, Yeosun, Zeynep Gürhan-Canli, and Norbert Schwarz. 2006. "The Effect of Corporate Social Responsibility (CSR) Activities on Companies with Bad Reputations." *Journal of Consumer Psychology* 16 (4): 377–90.
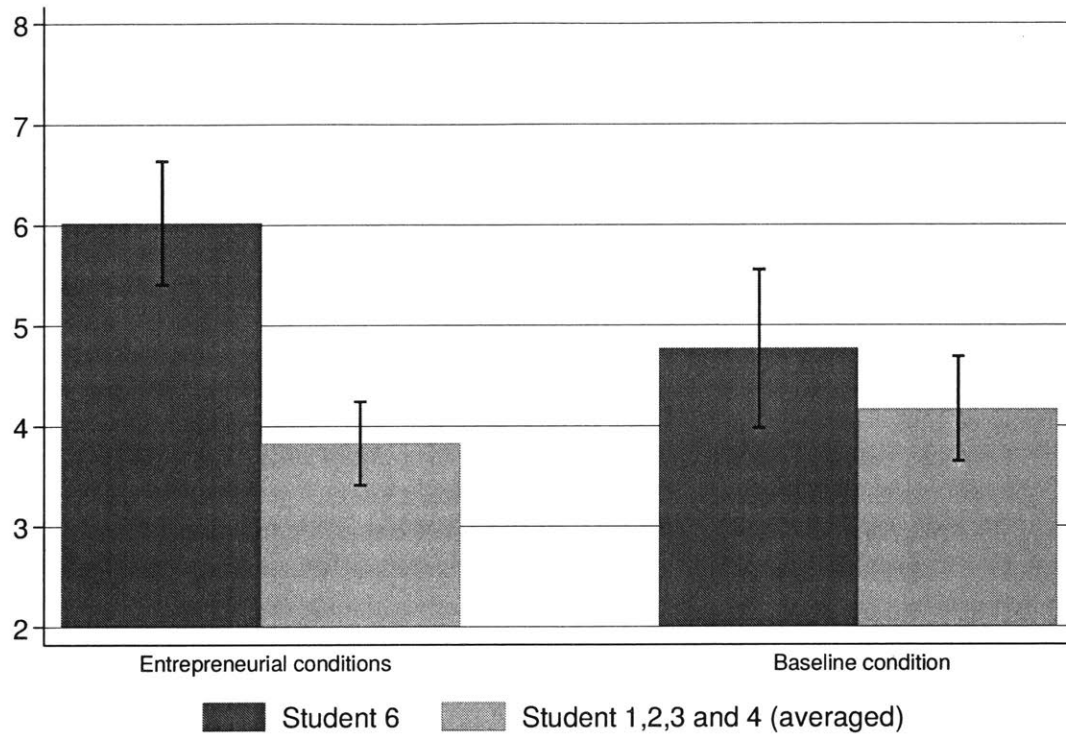
**FIGURES**

**Figure 1:** Between-Subject Comparison of perceived likelihood that Student 6 changed his response in Study 1
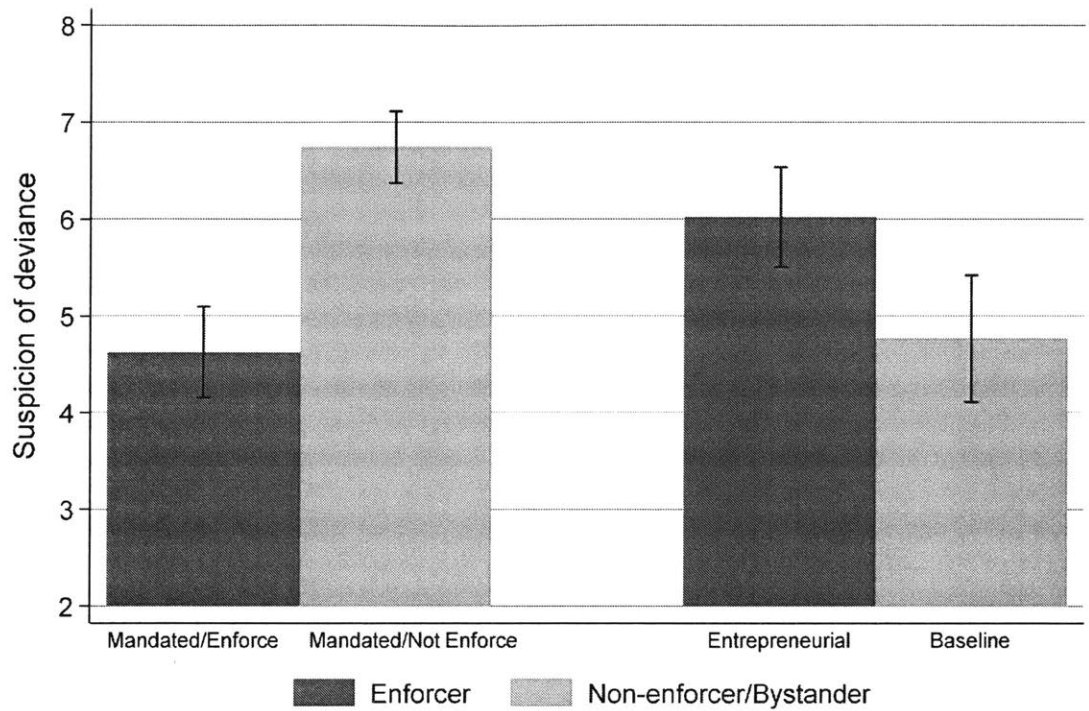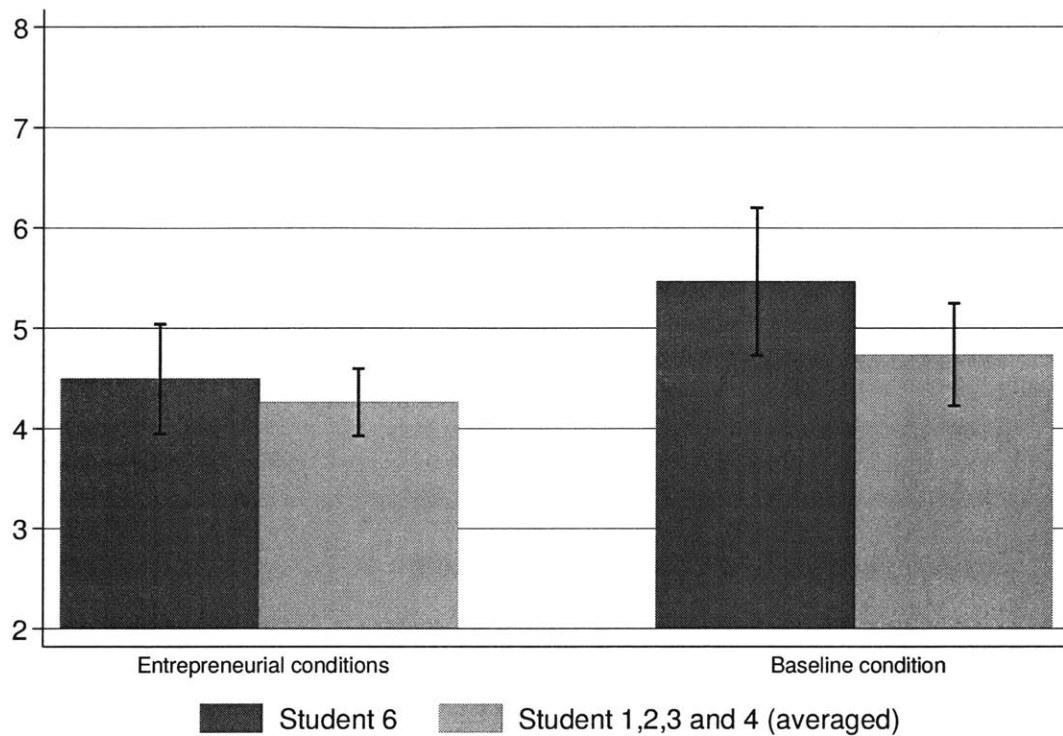


The error bars indicate 95% confidence interval. Possible value of the y-axis ranges from 1 (Very unlikely) to 10 (Very likely). The measure is the response to the question "How likely is it that Student X wrote the following statements on his anonymous questionnaire?: Student X wrote that "the college should ban alcohol on its campus.""

**Figure 2:** Within-Subject Comparisons of likelihood that the student in question changed his response in the *High-Accusability/Entrepreneurial* conditions and the *High-Accusability/Baseline* condition in Study 1
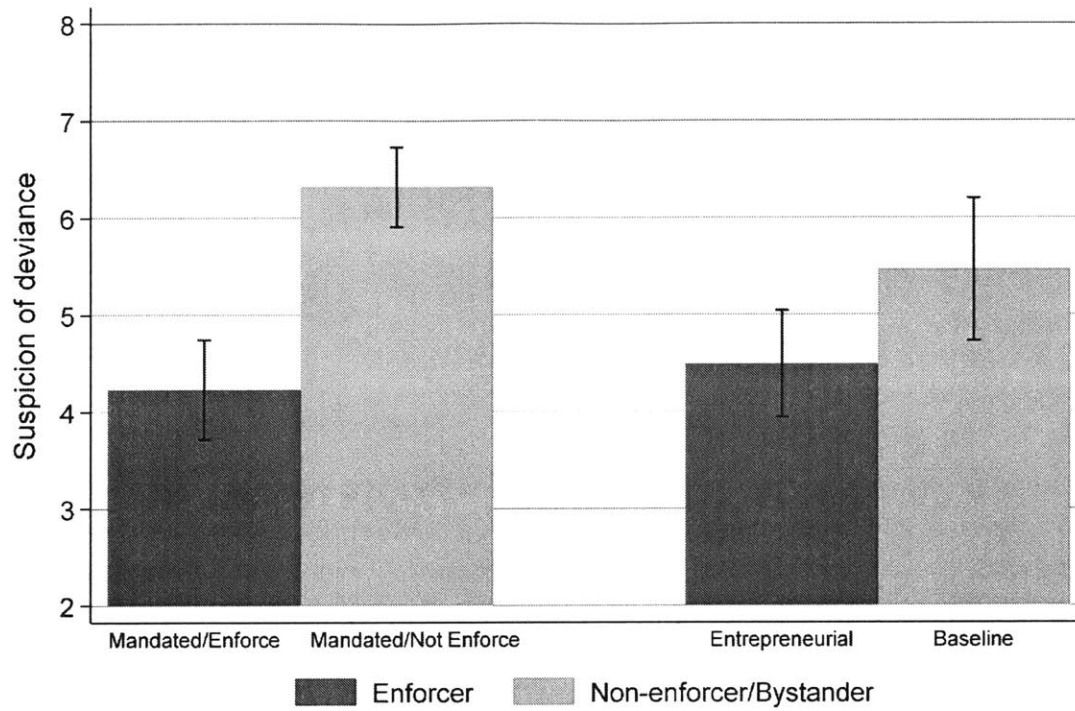


The error bars indicate 95% confidence interval. Possible value of the y-axis ranges from 1 (Very unlikely) to 10 (Very likely).  Separate measures for Students 1, 2, 3, and 4 were averaged in all the *High-Accusability/Entrepreneurial* conditions and the *High-Accusability/Baseline* condition (Cronbach's alpha=0.96).  The measure is the response to the following question: "How likely is it that Student X wrote the following statements on his anonymous questionnaire?: Student X wrote that "the college should ban alcohol on its campus.""

**Figure 3:** Within-Subject Comparisons of likelihood that the student in question changed his response in the *Low-Accusability/Entrepreneurial* conditions and the *Low-Accusability/Baseline* condition in Study 2



The error bars indicate 95% confidence interval. Possible value of the y-axis ranges from 1 (Very unlikely) to 10 (Very likely). Separate measures for Students 1, 2, 3, and 4 were averaged in all the *Low-Accusability/Entrepreneurial* conditions and the *Low-Accusability/Baseline* condition (Cronbach's alpha=0.88). The measure is the response to the following question: "How likely is it that Student X wrote the following statements on his anonymous questionnaire?: Student X wrote that "the college should ban alcohol on its campus.""""

**Figure 4:** Between-Subject Comparison of perceived likelihood that Student 6 changed his response in Study 2



The error bars indicate 95% confidence interval. Possible value of the y-axis ranges from 1 (Very unlikely) to 10 (Very likely). The measure is the response to the question "How likely is it that Student X wrote the following statements on his anonymous questionnaire?: Student X wrote that "the college should ban alcohol on its campus.""

**TABLES**

**Table 1:** Theoretical Predictions of When Norm Enforcers Seem Sincere or Suspicious

| Does enforcement occur in presence of mandate? | Mandate Present | Entrepreneurial enforcement |
|---|---|---|
| Low-accusability | *Sincere* | *Sincere* |
| High-accusability | *Sincere ("Illusion of Sincerity")* | *Suspicious* |

**Table 2:** Experimental conditions of Studies 1 and 2

**Study 1**

| High-accusability situation | Mandate to respond | No Mandate to respond |
| --- | --- | --- |
| **Enforcer** | *High-Accusability/ Mandated and Enforce* <br><br> (N=149) | *High-Accusability/ Entrepreneurial* <br><br> (N=150) |
| **Bystander (i.e., non-enforcer)** | *High-Accusability/ Mandated and Not Enforce* <br><br> (N=160) | *High-Accusability/ Baseline* <br><br> (N=50) |

**Study 2**

| Low-accusability situation | Mandate to respond | No Mandate to respond |
| --- | --- | --- |
| **Enforcer** | *Low-Accusability/ Mandated and Enforce* <br><br> (N=165) | *Low-Accusability/ Entrepreneurial* <br><br> (N=168) |
| **Bystander (i.e., non-enforcer)** | *Low-Accusability/ Mandated and Not Enforce* <br><br> (N=183) | *Low-Accusability/ Baseline* <br><br> (N=58) |

Note: Since the *Baseline* conditions do not have sub-conditions for different wording, the *Baseline* conditions have about a third of subjects of other conditions.

**Table 3:** Results of unbalanced ANOVA (Type II) assessing effect of *High-Accusability/Mandated and Enforce* (vs. *High-Accusability/Mandated and Not Enforce*) and different wordings in Study 1

| Variable | F | Df |
|---|---|---|
| *High-Accusability/Mandated and Enforce* | 35.33*** | 1 |
| Wording: *Simple* | 0.54 | 1 |
| Wording: *Stating-Principle* | 0.78 | 1 |
| *Mandated and Enforce* × *Simple* | 6.13* | 1 |
| *Mandated and Enforce* × *Stating-Principle* | 3.06 | 1 |
| Residuals | | 301 |

Note: Analysis for 2 *Mandated* conditions (*High-Accusability/Mandated and Enforce* and *High-Accusability/Mandated and Not Enforce*) × 3 wording sub-conditions (*Simple, Stating-Principle, and Activist*). The baseline condition is the *High-Accusability/Mandated and Not Enforce – Activist* sub-condition. The F-value for the *High-Accusability/Mandated and Enforce* variable represents the main effect of (i.e., variance explained by) enforcement (prompted by the mandate). The F-value for wording sub-conditions represents the added variance explained by each wording sub-condition of *High-Accusability/Mandated and Not Enforce* (i.e., how different is each *High-Accusability/Mandated and Not Enforce* wording sub-condition from the baseline condition). The F-value for interaction terms represents the added variance explained by each comparison of wording sub-conditions (i.e., how different is the main effect for the comparison between the two wording sub-conditions of *High-Accusability/Mandated and Enforce* and *High-Accusability/Mandated and Not Enforce*).

*p<0.05 (all tests are two-tailed)
**p<0.01
***p<0.001

**Table 4:** Results of unbalanced ANOVA (Type II) assessing effect of *High-Accusability/Entrepreneurial* (vs. *High-Accusability/Baseline*) and different wording sub-conditions in Study 1

| Variable | F | Df |
|---|---|---|
| *High-Accusability/Entrepreneurial* | 4.29* | 1 |
| *Entrepreneurial* × Wording: *Simple* | 0.06 | 1 |
| *Entrepreneurial* × Wording: *Principle* | 0.39 | 1 |
| Residuals | | 183 |

Note: The baseline condition is the *High-Accusability/Baseline* condition. Because the *High-Accusability/Baseline* condition does not vary in wording, there are no sub-conditions corresponding to any sub-conditions in the *High-Accusability/Baseline* condition. Since the *Activist* wording is the most plausible, the baseline wording condition is the *High-Accusability/Entrepreneurial – Activist* sub-condition. The F-value for the *High-Accusability/Entrepreneurial* variable represents the main effect of (i.e., variance explained by) entrepreneurial enforcement in the *Activist* sub-condition. The F-value for the other two variables (i.e., interaction terms) is for the added variance explained by each wording sub-condition.

*p<0.05 (all tests are two-tailed)

**p<0.01

***p<0.001

**Table 5:** Results of unbalanced ANOVA (Type II) assessing effect of *High Accusability* (i.e., Study 1) vs. *Low Accusability* (i.e., Study 2) and different wording sub-conditions within *Entrepreneurial* conditions

| Variable | F | Df |
|---|---|---|
| *High Accusability* (Study 1) | 13.49*** | 1 |
| Wording: *Simple* | 2.63 | 1 |
| Wording: *Stating-Principle* | 0.00 | 1 |
| *High Accusability* × *Simple* | 3.34 | 1 |
| *High Accusability* × *Stating-Principle* | 0.57 | 1 |
| Residuals | | 302 |

Note: Analysis for 2 *Entrepreneurial* conditions (one from Study 1 [*High-Accusability*] and one from Study 2 [*Low-Accusability*]) × 3 wording sub-conditions (*Simple, Stating-Principle, and Activist*). The baseline condition is the *Low-Accusability/Entrepreneurial – Activist* sub-condition from Study 2. The F-value for the '*High Accusability*' variable represents the main effect of (i.e., variance explained by) increased accusability in Study 1. The F-value for wording sub-conditions represents the added variance explained by each wording sub-condition of *Low-Accusability/Entrepreneurial* in Study 2. The F-value for interaction terms represents the added variance explained by each wording sub-condition of *High-Accusability/Entrepreneurial* in Study 1.
*p<0.05 (all tests are two-tailed)
**p<0.01
***p<0.001

**Table 6:** Results of unbalanced ANOVA (Type II) assessing effect of *Low-Accusability/Entrepreneurial* (vs. *Baseline*) and different wording sub-conditions in Study 2

| Variable | F | Df |
|---|---|---|
| *Low-Accusability/Entrepreneurial* | 6.61* | 1 |
| *Entrepreneurial* × Wording: *Simple* | 6.79** | 1 |
| *Entrepreneurial* × Wording: *Stating-Principle* | 0.25 | 1 |
| Residuals | | 222 |

Note: The baseline condition is the *Low-Accusability/Baseline* condition. Because the *Low-Accusability/Baseline* condition does not vary in wording, there are no sub-conditions corresponding to any sub-conditions in the *Low-Accusability/Baseline* condition. Since the *Activist* wording is the most plausible, the baseline condition is the *Low-Accusability/Entrepreneurial – Activist* sub-condition. The F-value for the *Low-Accusability/ Entrepreneurial* variable represents the main effect of (i.e., variance explained by) entrepreneurial enforcement. The F-value for the other two variables (i.e., interaction terms) is for the added variance explained by each wording sub-condition.

*$p < 0.05$ (all tests are two-tailed)

**$p < 0.01$

***$p < 0.001$

**Table 7:** Results of unbalanced ANOVA (Type II) assessing effect of *Low-Accusability/Mandated and Enforce* (vs. *Low-Accusability/Mandated and Not Enforce*) and different wording sub-conditions in Study 3

| Variable | F | Df |
|---|---|---|
| *Mandated and Enforce* | 36.33*** | 1 |
| Wording: *Simple* | 0.31 | 1 |
| Wording: *Stating-Principle* | 1.57 | 1 |
| *Mandated and Enforce* × *Simple* | 2.53 | 1 |
| *Mandated and Enforce* × *Stating-Principle* | 3.92* | 1 |
| Residuals | | 342 |

Note: Analysis for 2 *Mandated* conditions (*Mandated and Enforce* and *Mandated and Not Enforce*) × 3 wording sub-conditions (*Simple, Stating-Principle, and Activist*). The baseline condition is the *Mandated and Not Enforce – Activist* sub-condition. The F-value for the *Mandated and Enforce* variable represents the main effect of (i.e., variance explained by) enforcement (prompted by the mandate). The F-value for wording sub-conditions represents the added variance explained by each wording sub-condition of *Mandated and Not Enforce* (i.e., how different is each *Mandated and Not Enforce* wording sub-condition from the baseline condition). The F-value for interaction terms represents the added variance explained by each comparison of wording sub-conditions (i.e., how different is the main effect for the comparison between the two wording sub-conditions of *Mandated and Enforce* and *Mandated and Not Enforce*).

+p<0.1 (all tests are two-tailed)
*p<0.05
**p<0.01
***p<0.001

## METHODOLOGICAL APPENDIX

As reviewed in the main text, the strongest evidence that enforcement can create an "illusion of sincerity" comes from Study 3 of Willer et al. (2009). Here we lay out the major methodological differences between that study and our studies and explain why we developed a related but distinct experimental paradigm in order to test our theory.

In Study 3 of Willer et al. (2009), subjects learned (i) that three students had read an unintelligible yet purportedly scholarly text; (ii) that the three students had each given private evaluations of the text; (iii) that the students then had a discussion in which two expressed a favorable opinion of the text while one dissented with a negative view; and (iv) that one of the students who publicly expressed a favorable opinion was asked to rate the other students' performance. The premise of this experiment is that subjects recognized that the student had an ulterior motive to give a low rating to the public deviant. Yet despite such recognition, subjects were more likely to believe that this student had privately expressed a favorable view when he gave the public deviant a negative evaluation than when he gave the public deviant a neutral evaluation. Based on these results, Willer et al. suggest that audiences do not perceive that the negative evaluation t is driven by the ulterior motive, perhaps due to cognitive limitations.[43] Our theory suggests an alternative—i.e., that suspicion is low when assessments of others' actions occur in response to a mandate.

While it might seem natural to test this idea by replicating and extending Study 3 of Willer et al., there are several aspects in Study 3 of Willer et al. that make it an inappropriate setting to test when ulterior motives become salient in acts of norm enforcement. We thus

---

[43] For more details on their design and results, see: Willer et al. (2009:476-481).

modified those aspects, but were still able to reproduce the key results of Willer at al. (in the

*Mandated* conditions of Study 1).


*Modifications*

The most important change is that conformity in our studies means adhering to a norm

rather than performing in accordance with a group performance standard. As reviewed above,

the fictive participants in the vignettes presented in Willer at al. Study 3 evaluated the quality of

a scholarly text and then evaluated one another's assessment of that text. As in the classic Asch

(1951) experiments, the group majority sets a performance standard. Those in the minority

appear to be less *capable* than those in the majority. But this is subtly different from conformity

with a group norm.[44] Noncompliance with a norm is not regarded as problematic if the

individual is recognized as incapable of adhering to the norm (e.g., van Maanen 1973). The issue

instead pertains to individuals who could choose to join the majority in conforming to the norm

but decide not to do so. This indicates problematic *commitment*.[45] Based on these

considerations, we alter the scenario from one in which individuals judge the quality of a piece

of work to one where individuals assess the appropriate norms for their community, as shown in

the main text.[46]

---

[44] Phillips, Turco, and Zuckerman (2013) review this issue.

[45] This contrast may not be so sharp in some contexts. For instance, one may be able to successfully appear committed to the norm even after changing her belief to match those of others, if she can produce a new principle behind supporting the norm. Therefore, as reviewed above, we varied how the enforcer and the non-enforcer articulate her belief with different wording sub-conditions, and we tested the effect of enforcement through different plausible scenarios.

[46] Underlying this assessment are beliefs about the welfare of the community and how group members ought to behave to increase its welfare. By contrast, although promoting a performance standard might shore up one's own status within the group, it is harder to draw the link to group welfare. Note however that some situations of normative judgment involve a mix of individual versus group-oriented justifications. For instance, consider norms against theft of intellectual property (e.g., Di Stefano, King, and Verona 2015). Enforcing such a norm is generally motivated by an attempt to protect one's interests or those of one's allies, but enforcement is justified in terms of overall group welfare: we should all

This distinction is also crucial in sharpening when and why audiences recognize strategic reasons for enforcement. Willer et al. present their study as addressing the question of why actors would enforce an "unpopular norm." As noted, their case is actually not one of norms but of performance standards. In addition, the label "unpopular" is applied based on the premise that since the scholarly text is nonsensical, each subject has a private experience of the performance that is below that of the majority's public assessment. The ulterior motive for enforcement therefore derives from insecurity about one's capability: *if everyone else thinks the text is compelling, maybe I'm wrong and will be called a fool for missing it?* However, it is not clear how this insecurity about one's *capability* translates to the context of norms, especially when multiple norms compete with one another: there is no *a priori* reason for individuals to guess where others stand (e.g., supporting or opposing allowing alcohol on campus). In the context of norms, therefore, the strategic reason to enforce comes from *insecurity about one's commitment or status in the group*; and the group's commitment is likely inferred from the group majority's opinion (e.g., Prentice and Miller 1993; cf., Kim 2017). By investigating a setting of normative debate, we sharpen when and why there might be strategic reasons for appearing committed—i.e., where the group majority seems to endorse a norm.

Finally, it is useful to clarify why our studies involve a debate between competing principles—a libertarian position that is opposed to the alcohol ban and a health/welfare-based position that is in favor. One reason is that such cases are quite common (Duncan 2015). Our design is reflective of the real-world examples where there are often coherent principles for both sides (e.g., being a Democrat or Republican on college campus; for or against abolishment of

---

protect my rights because your rights will be threatened next. It would seem that suspicions of ulterior motive are even more salient in such conditions (cf., Reilly 2016), thus making our setting a conservative test.

capital punishment). A second reason is methodological. In particular, if there is not a coherent alternative to the majority opinion, subjects may think that the (fictitious) students are conforming and enforcing simply because the opposing argument is weak; that is, it becomes a matter of performance rather than that of commitment.

Beyond these main differences between our studies and Study 3 of Willer et al., there are also more minor differences:

*Number of Students in the Group:* Our study's fictitious student group consisted of six students, instead of three students as in Study 3 of Willer et al., because it is questionable whether there is any social pressure when the majority consists of two out of a group of three. After all, if the third switches, the majority moves with her: thus social pressure seems rather weak.

*Accidental Revelation of a Deviant:* Whereas the deviant in Study 3 of Willer et al. revealed himself as a deviant (by declaring that the scholarly text was unintelligible), we designed the story so that the deviant is revealed by accident. If someone is willing to voluntarily reveal himself as a deviant as in Study 3 of Willer et al., it becomes questionable whether the norm exerts social pressure on members of the group. By revealing the deviance accidentally, it remains a viable interpretation that being supportive of banning alcohol is not simply a minority opinion, but an embattled one. However, as in Study 3 of Willer et al., it obviously remains unknown who might be a closeted deviant.

*Revelation of a Closeted Deviant:* our experiments create situations where an unknown deviant is known to hide among those who apparently are genuine in endorsing the norm. To recall the case of the Chinese Cultural Revolution, the key opening event was Mao's declaration that there were traitors in their midst. This is also common in more mundane situations, where

deviance is witnessed (e.g., a failure of collective action) but it is less clear who the deviant is

(e.g., Alchian and Demsetz 1972). In a setting we present as a benign academic exercise,

audiences would have no reason to suspect ulterior motives of anyone without hearing about the

(hidden) presence of a deviant.

**Appendix Table 1:** Wordings for each condition's peer evaluation

| | Simple | Stating-principle | Activist |
|---|---|---|---|
| *Mandated and Enforce*[47] | "I obviously think that opinions of Students 1, 2, 3, and 4 are very valid. But I think Student 5's opinion—that alcohol should be banned—is not valid at all." | "I obviously think that opinions of Students 1, 2, 3, and 4 are very valid. But I think Student 5's opinion—that alcohol should be banned—is not valid at all. Student 5 does not appreciate the importance of individual freedom. This is not a value on which we can compromise." | "I obviously think that opinions of Students 1, 2, 3, and 4 are very valid. But I think Student 5's opinion—that alcohol should be banned—is not valid at all. Student 5 does not appreciate the importance of individual freedom. This is not a value on which we can compromise. Actually, I really want to make sure that the college doesn't ban alcohol! Let's get together in my dorm room after the session to talk about how to stop the college from banning alcohol." |
| *Mandated and Not Enforce* | "I think that the opinions of Students 1, 2, 3, 4, and 5 are all equally valid." | "I obviously think that the opinions of Students 1, 2, 3, and 4 are quite valid. While I ultimately disagree with him, I also think that Student 5's opinions are quite valid too. Don't get me wrong—I think that we need to make sure that we uphold the principle of individual freedom. But let's also make | "I obviously think that the opinions of Students 1, 2, 3, and 4 are quite valid. I also think that Student 5's opinions are quite valid too. It is important to make sure that no underage students feel pressure to drink, and I agree it's important to keep our mind and body healthy." |

[47] Since Student 6 in the *Mandated and Enforce* conditions was asked to comment on all the other students as well as Student 5, Student 6 in the *Mandated and Enforce* conditions expresses his agreement with Students 1, 2, 3, and 4 as well as disagreement with Student 5. It is unrealistic, however, to add this to Student 6's wording in the *Entrepreneurial* conditions, since Student 6 has no reason to comment on those who agreed with Student 6 in the first place.

sure that we keep listening to people like Student 5."

| | | | |
|---|---|---|---|
| *Entrepreneurial* | "I just want to say that I think Student 5's opinion—that alcohol should be banned—is not valid at all." | "I just want to say that I think Student 5's opinion—that alcohol should be banned—is not valid at all. Student 5 does not appreciate the importance of individual freedom. This is not a value on which we can compromise." | "I just want to say that I think Student 5's opinion—that alcohol should be banned—is not valid at all. Student 5 does not appreciate the importance of individual freedom. This is not a value on which we can compromise. Actually, I really want to make sure that the college doesn't ban alcohol! Let's get together in my dorm room after the session to talk about how to stop the college from banning alcohol." |
| *Baseline* | No one says anything after the administrator announces that there is someone who changed his opinion. | | |