ACADEMY OF Management Journal

Diversity Thresholds: How Social Norms, Visibility, and Scrutiny Relate to Group Composition

Journal:	*Academy of Management Journal*
Manuscript ID	AMJ-2017-0440.R2
Manuscript Type:	Revision
Keywords:	Gender and Diversity in Organizations < Topic Areas, Decision making (micro perspectives) < Behavior < Organizational Behavior < Topic Areas, Impression management < Organizational Behavior < Topic Areas, Composition/diversity < Group/team characterisitics < Organizational Behavior < Topic Areas, Board of directors < Upper Echelons/Corporate Governance < Business Policy and Strategy < Topic Areas
Abstract:	Across a field study and four experiments, we examine how social norms and scrutiny affect decisions about adding members of underrepresented populations (e.g., women, racial minorities) to groups. When groups are scrutinized, we theorize that decision makers strive to match the diversity observed in peer groups due to impression management concerns, thereby conforming to the descriptive social norm. We examine this first in the context of U.S. corporate boards where firms face pressure to increase gender diversity. Analyses of S&P 1500 boards reveal that significantly more boards include exactly two women (the descriptive social norm) than would be expected by chance. This overrepresentation of two-women boards–a phenomenon we call "twokenism"–is more pronounced among more visible companies, consistent with our theorizing around impression management and scrutiny. Experimental data corroborate these findings and provide support for our theoretical mechanism: decision makers are discontinuously less likely to add a woman to a board once it includes two women (the social norm), and decision makers' likelihood of adding a woman or minority to a group is influenced by the descriptive social norms and scrutiny faced. Together, these findings provide a new perspective on the persistent underrepresentation of women and minorities in organizations.

Diversity Thresholds: How Social Norms, Visibility, and Scrutiny Relate to Group Composition

Edward H. Chang
University of Pennsylvania
changed@wharton.upenn.edu

Katherine L. Milkman
University of Pennsylvania
kmilkman@wharton.upenn.edu

Dolly Chugh
New York University
dchugh@stern.nyu.edu

Modupe Akinola
Columbia University
makinola@columbia.edu

Diversity Thresholds: How Social Norms, Visibility, and Scrutiny Relate to Group Composition

ABSTRACT

Across a field study and four experiments, we examine how social norms and scrutiny affect decisions about adding members of underrepresented populations (e.g., women, racial minorities) to groups. When groups are scrutinized, we theorize that decision makers strive to match the diversity observed in peer groups due to impression management concerns, thereby conforming to the descriptive social norm. We examine this first in the context of U.S. corporate boards where firms face pressure to increase gender diversity. Analyses of S&P 1500 boards reveal that significantly more boards include exactly two women (the descriptive social norm) than would be expected by chance. This overrepresentation of two-women boards–a phenomenon we call "twokenism"–is more pronounced among more visible companies, consistent with our theorizing around impression management and scrutiny. Experimental data corroborate these findings and provide support for our theoretical mechanism: decision makers are discontinuously less likely to add a woman to a board once it includes two women (the social norm), and decision makers' likelihood of adding a woman or minority to a group is influenced by the descriptive social norms and scrutiny faced. Together, these findings provide a new perspective on the persistent underrepresentation of women and minorities in organizations.

In recent years, many groups have faced negative scrutiny for their lack of diversity. For instance, the Academy of Motion Picture Arts and Sciences faced backlash in 2015 and in 2016 when all twenty actors nominated for Academy Awards in the lead and supporting acting categories were white. This sparked an #OscarsSoWhite meme and a plan to double female and minority membership in the Academy by 2020 (Ryan, 2016). When Twitter made an initial public offering with no women on its board of directors in 2013, the company faced an outpouring of negative media attention, with numerous outlets claiming that the lack of gender diversity would cause problems for the company (Merchant, 2013; Miller, 2013). And when Donald Trump announced the members of his presidential cabinet in 2017, the New York Times ran a front-page story tallying the women and racial minorities Trump's cabinet included and comparing its (lack of) diversity to all other modern U.S. administrations (Lee, 2017). These examples illustrate that when groups lack diversity, negative scrutiny–or critical attention paid to particular behaviors (Sutton & Galunic, 1996)–can ensue.

Little is known, however, about when a group's diversity will be judged negatively or how groups will respond to the possibility of negative scrutiny regarding their diversity. While scholarship has established that diversity is not perceived objectively, or equivalently, by all observers and in all contexts (Unzueta & Binning, 2010, 2012; Unzueta, Knowles, & Ho, 2012), it remains ambiguous as to when group members and those perceiving groups judge a group's diversity to be so insufficient as to warrant action or attention. Further, although past work has established that organizations respond to reputational threats such as social movement boycotts (King, 2008; McDonnell & King, 2013), it is unclear how those responsible for group composition may behave when facing the threat of repercussions for displaying insufficient diversity. In this paper, we address these questions by analyzing a decade of data on the composition of U.S. corporate boards in the S&P 1500 and by conducting a series of supplemental experiments.

We propose that, to avoid facing negative scrutiny, those responsible for forming groups may seek safety in numbers by looking to the average behavior of others when setting implicit or explicit goals about the diversity of groups. Descriptive social norms–defined as the average observed behavior of individuals or groups in a population (Prentice & Miller, 1993)–have been shown to serve as reference points for behavior in a variety of contexts, setting expectations about what is appropriate and effective (Coffman, Featherstone, & Kessler, 2014; Goldstein, Cialdini, & Griskevicius, 2008; Nolan, Schultz, Cialdini, Goldstein, & Griskevicius, 2008), particularly in situations where appropriate behavior is ambiguous or uncertain (Festinger, 1954; Sherif, 1936). Decision makers and firms may thus look to relevant others to understand what the descriptive social norms for diversity are, and they may then imitate these levels of diversity, both because of the reputational threat associated with negative scrutiny and because of

uncertainty about what adequate diversity entails (DiMaggio & Powell, 1983). This behavior should be even more prevalent among highly visible groups or organizations because the negative consequences of failing to conform can be greater for high-profile groups (Gardberg & Fombrun, 2006). The actions of highly visible groups are more likely to be scrutinized in the first place (Chiu & Sharfman, 2011), and organizations generally respond more strongly to more visible threats (King, 2008).

We combine our theorizing about descriptive social norms, scrutiny, and visibility with past research on goal setting to make a novel prediction. Specifically, we predict that individuals responsible for group compositions will respond to pressures to diversify in a similar fashion, leading to an overabundance of groups with identical levels of diversity. Past research has shown goals–like the goal to match the diversity of peer groups–are often highly motivating (Locke & Latham, 2002), but individuals relax efforts to achieve desirable outcomes after reaching salient goal thresholds in many settings (Heath, Larrick, & Wu, 1999). This relaxing of effort has been shown to lead goal-seekers' performance to cluster around salient goal thresholds (Pope & Simonsohn, 2011). We predict this tendency will lead scrutinized groups to cluster around the social norm for diversity set by their peers. In other words, rather than continuing to increase diversity in response to external pressures (e.g., the threat of negative scrutiny), those with the power to shape group diversity should be less likely to increase the diversity of a group once the group has reached the descriptive social norm for diversity set by peers. This behavior will lead to improbably homogeneous diversity levels across groups.

We test our theorizing first in the context of U.S. corporate boards, a setting where firms face negative scrutiny for failing to include adequate gender diversity (Merchant, 2013; Miller, 2013). Analyses of S&P 1500 boards reveal that significantly more boards include exactly two

women (the descriptive social norm) than would be expected by chance, supporting our prediction that groups will respond to pressures to diversify in a similar fashion, leading to an overabundance of groups with identical levels of diversity at the descriptive social norm. This overrepresentation of two-women boards is more pronounced among more visible companies, consistent with our theorizing around impression management and scrutiny. In additional studies, we experimentally manipulate descriptive social norms, scrutiny, and visibility to show that each of these influences group diversity decisions as our theory predicts in groups besides corporate boards and when we examine social categories besides gender.

Our work provides a more complete understanding of diversity-related hiring decisions, telling us when women and racial minorities will be particularly attractive candidates for inclusion in groups and when groups will reduce their efforts to increase diversity. Further, rather than focusing only on individual-level or firm-level explanations for why women and racial minorities may or may not be added to groups, we highlight how external entities such as peers (who help shape descriptive social norms) and outsider scrutiny can shape group diversity decisions. By illuminating these critical factors that influence group diversity decisions, we provide theoretical guidance about potential new ways to improve diversity in organizations and practical guidance to help predict what levels of diversity we might expect to see in different contexts. Our research suggests that it may be helpful to increase scrutiny around diversity decisions and attempt to make other social norms besides descriptive social norms salient to decision makers in order to increase the number of women and racial minorities selected into groups.

THEORY AND HYPOTHESES

Descriptive Social Norms

Descriptive social norms–defined as the average observed behavior of individuals or groups in a population (Prentice & Miller, 1993)–exert a potent influence on decisions. According to past research, descriptive social norms influence the behavior of individuals and groups for two primary reasons. First, they establish what is socially acceptable. Because following the norm means avoiding outlier status, individuals and groups can feel reassured that if existing norms are followed, social ostracism will not ensue (Schultz, Nolan, Cialdini, Goldstein, & Griskevicius, 2007). By following a descriptive social norm, individuals and groups essentially insulate themselves from the risk of being singled out because they are–by definition–doing what many of their peers are doing. Individuals, groups, and organizations that negatively deviate from any descriptive norm are much more likely to be singled out and face negative consequences (Ahmadjian & Robinson, 2001; Zavyalova, Pfarrer, Reger, & Shapiro, 2012).

Second, descriptive social norms contain information about what behaviors are likely to be effective or adaptive (Cialdini, 2007). If the majority of others have elected to partake in a specific action or behavior (making it the descriptive social norm), then that signals that the norm may be a wise course of action (e.g., if everyone else is using this brand of soap, it must be a good brand of soap to use). This social information is even more important when the appropriate behavior is unclear or when situations are ambiguous or uncertain, as extant research has shown that social norms affect behavior to a greater degree in such settings (Festinger, 1954; Sherif, 1936). In effect, descriptive social norms can function as heuristics for decision making, providing a guide for appropriate or wise behavior in a wide range of situations.

Together, both by conveying what is appropriate and likely to be effective, descriptive social norms produce powerful effects on judgments and decisions (Cialdini, 2003; Cialdini, Reno, & Kallgren, 1990). A large body of empirical evidence has shown that descriptive social

norms serve as salient reference points for behavior in many contexts, ranging from energy consumption to job acceptance decisions (Coffman et al., 2014; Goldstein et al., 2008; Nolan et al., 2008). We propose that descriptive social norms should influence decisions made about group diversity just as they influence decisions in other contexts. Past research on scrutiny and impression management illuminates why those responsible for decisions influencing group diversity may feel pressure to follow descriptive social norms.

How Scrutiny of Group Diversity May Drive Conformity to Descriptive Social Norms

Scrutiny refers to obtrusive and critical attention paid to particular behaviors (Sutton & Galunic, 1996), and scrutiny can come from a variety of sources. For example, the media is one common source of scrutiny capable of influencing an organization's reputation and value and shaping others' perceptions of its legitimacy. Naturally, organizations compete to receive positive and avoid negative media exposure (Fombrun, 1996; Fombrun & Shanley, 1990; Pollock & Rindova, 2003). Scrutiny can also come from other sources such as shareholders (e.g., institutional investors placing pressure on firms to engage in socially responsible behaviors) and policy makers (e.g., through regulations and the imposition of rewards or penalties for certain behaviors; Aguilera, Rupp, Williams, & Ganapathi, 2007; Campbell, 2007). The public also often directly scrutinizes organizations, mobilizing in ways that may draw wanted or unwanted attention to particular behaviors (e.g., through social movement boycotts; McDonnell & King, 2013).

In general, groups and organizations have strong incentives to avoid negative scrutiny. Negative scrutiny can be detrimental for reputation and legitimacy (Desai, 2011), so in order to avoid negative scrutiny, groups frequently attempt to manage impressions around scrutinized behaviors (Bolino, Kacmar, Turnley, & Gilstrap, 2008; Elsbach, Sutton, & Principe, 1998).

Impression management describes attempts by groups or organizations to positively shape how they are perceived (Elsbach & Sutton, 1992), and it may occur even in anticipation of the possibility of negative events. For example, Elsbach et al. (1998) have documented how hospitals use anticipatory impression management tactics in order to prevent potential negative scrutiny.

In recent years, scrutiny has increased surrounding the diversity of groups. For example, the media has scrutinized companies for insufficient gender diversity on their boards of directors (Merchant, 2013; Miller, 2013); presidents for insufficient race and gender diversity in their cabinets (Lee, 2017) and their U.S. Supreme Court nominees (Totenberg, 2016); and the Academy of Motion Picture Arts and Sciences for insufficient racial diversity among their Oscar nominees (Buckley, 2016; Ryan, 2016). Importantly, scrutiny is often applied selectively: rather than simultaneously emphasizing racial, gender, and socio-economic diversity, for instance, scrutiny often focuses more narrowly on a single dimension of diversity. For example, while groups such as corporate boards have faced considerable negative scrutiny for a lack of gender diversity, there has been far less attention to their lack of racial diversity.

Scrutiny surrounding diversity naturally motivates impression management concerns. An important question, then, is how decision makers who shape the composition of high profile groups within organizations may seek to manage diversity in order to avoid negative scrutiny. We propose that past research on descriptive social norms provides key insights. If groups or organizations are motivated to avoid negative scrutiny, then following the descriptive social norm for diversity essentially ensures that they will not be singled out for inadequate diversity. Further, because it is often unclear what an "objective" benchmark for strong performance should be in the context of decisions around diversity (Bell & Hartmann, 2007; Shemla, Meyer,

Greer, & Jehn, 2014; Unzueta et al., 2012), descriptive social norms should be particularly informative in guiding behavior around diversity. Thus, groups and organizations (and the decision makers responsible for their composition) may treat the descriptive social norm for diversity as a goal for impression management reasons.

The Implications of Descriptive Social Norms as Diversity Goals

Past research on goal setting offers insight into what will happen when those who shape group composition share the same explicit or implicit goal. Goals serve as reference points, causing individuals to expend considerable effort in the hopes of achieving an unmet goal and then to relax their efforts after achieving it (Heath et al., 1999; Locke & Latham, 2002). This has been shown to lead to performance clustering around salient goal thresholds in numerous contexts. For instance, professional baseball players finish seasons disproportionately often with a batting average just above .300 (a salient threshold widely believed to separate good hitters from great ones; Moskowitz & Wertheim, 2011; Pope & Simonsohn, 2011), and marathon runners finish races disproportionately often in the minute right before salient, round number thresholds (e.g., the minute just under three hours; Allen et al., 2016). We therefore expect to observe an excess mass or clustering of groups at (or just above) the descriptive social norm for diversity.[1]

Hypothesis 1a. Groups' diversity levels will cluster at (or just above) the descriptive social norm set by peers for diversity.

While Hypothesis 1a pertains to group composition, group composition is the result of decisions regarding which members to add to a group. If reaching the descriptive social norm for

[1] Because descriptive social norms are averages, they are rarely whole numbers (e.g., the average number of women per board was 1.36 women in the S&P 1500 in 2013). Since groups cannot have fractional numbers of women or racial minorities, we expect clustering at "or just above" the descriptive social norm (i.e. at the smallest whole number above the descriptive social norm).

diversity is a goal of those who shape group compositions, then efforts to increase group diversity (in the form of adding underrepresented group members) should decline precipitously once the descriptive social norm for diversity is achieved. Empirically, this relaxing of effort after reaching a goal threshold has been observed in several contexts. In the context of baseball, as just mentioned, batters and their teams reduce their at bat appearances near the end of the season once they have exceeded the salient .300 batting average threshold that separates good hitters from great ones (Pope & Simonsohn, 2011). In the context of SAT scores, students are disproportionately less likely to retake the SAT once they surpass a salient threshold such as a score of 1000 (the average score set by the College Board and a salient round number; Pope & Simonsohn, 2011). In our context of diversity and group composition decisions, we predict that groups are less likely to increase their diversity once they have already reached the descriptive social norm for diversity established by peers.

Hypothesis 1b. Groups (and the individuals who shape their composition) will add new members from underrepresented populations at a lower rate once they have surpassed the pertinent descriptive social norm for diversity.

Importantly, we only expect descriptive social norms to serve as goals when it comes to scrutinized dimensions of diversity. Without any scrutiny on a given dimension of diversity, there should be no impression management motives and thus no desire to follow the descriptive social norm. For example, we would expect to find support for Hypotheses 1a and 1b when it comes to gender diversity in settings where inadequate gender diversity has been scrutinized (e.g., on corporate boards) but not in settings where gender diversity has not been scrutinized. Thus, we propose that scrutiny (or the threat of negative scrutiny) is required in order to produce our hypothesized clustering and threshold effects.

Hypothesis 2. Scrutiny moderates the effects of descriptive social norms on group diversity decisions. Specifically, descriptive social norms will only influence group diversity decisions and outcomes when scrutiny is present along a given diversity dimension.

The Moderating Role of Visibility

If groups and organizations manage impressions around diversity to avoid negative scrutiny, this tendency should be more pronounced among more visible groups and organizations. We follow past research and use the term "visibility" to describe how much attention individuals, groups, or organizations typically receive (Chiu & Sharfman, 2011), regardless of why they are receiving this attention (as opposed to our use of the term "scrutiny", which refers to attention paid to a particular behavior such as a group's gender diversity). When firms are more visible (e.g., because they operate in more visible industries or because they have higher overall media exposure), they face greater external pressures to engage in legitimacy-seeking behaviors (Gardberg & Fombrun, 2006) and are also more likely to engage in legitimacy-enhancing behaviors like corporate social performance initiatives (Chiu & Sharfman, 2011). For example, firms respond more to boycotts when they receive more media attention (King, 2008), and firms engage in more prosocial activities when boycotts are more threatening because of increased media attention (McDonnell & King, 2013). Past research has shown that conforming to descriptive social norms (i.e. mimicking the behavior of peer firms) is one way to enhance legitimacy (DiMaggio & Powell, 1983), suggesting that descriptive social norms should influence the diversity of groups along scrutinized diversity dimensions to a greater degree when those groups are more visible. Further, the actions of more visible firms receive more attention, which can magnify the negative consequences of failing to conform to social norms.

Past research on individual judgment and decision making makes similar predictions regarding the effects of visibility on conformity to descriptive social norms. Social norms influence behavior to a greater degree when individuals and their behaviors are more visible (Cialdini & Trost, 1998). In particular, individuals tend to look to social norms to guide their behavior most frequently when the behavior in question is public or observable (Cialdini, Kallgren, & Reno, 1991; Cialdini et al., 1990; Kallgren, Reno, & Cialdini, 2000; Shaffer, 1983). For example, studies have found that monitoring employees can improve conformity to ethical norms in the context of employee theft (Pierce, Snow, & McAfee, 2015), monitoring can improve conformity to hand hygiene norms in hospitals (Staats, Dai, Hofmann, & Milkman, 2016), and being in a public setting (as opposed to a private setting) can make women more likely to conform to gender norms regarding assertiveness (Swim & Hyers, 1999). On an individual level, we would thus expect more conformity to descriptive social norms when outcomes are more visible. Thus, research and theorizing on both individuals and firms suggests more visible groups should be more likely to conform to social norms around diversity along scrutinized diversity dimensions.

Hypothesis 3. Visibility moderates the effects of descriptive social norms on group diversity decisions along scrutinized diversity dimensions. Specifically, more visible groups will be more likely to follow the descriptive social norm for diversity along scrutinized diversity dimensions than less visible groups.

OVERVIEW OF STUDIES

The remainder of this paper proceeds as follows. We begin by examining our hypotheses in the field, exploring whether they make accurate predictions about the composition and evolution of U.S. corporate boards. In Study 1A, we present analyses of S&P 1500 board

composition data from 2013 that test for excess clustering of corporate boards at the descriptive social norm for gender diversity (Hypothesis 1a). We also examine whether this pattern is more extreme among more visible companies (Hypothesis 3). In Study 1B, we present analyses of board member additions to determine whether boards are discontinuously less likely to add female directors once they have reached the descriptive social norm for gender diversity (Hypothesis 1b). In Study 1C, we run an online experiment to test for evidence of the same pattern of discontinuities in board member selection found in the field in Study 1B in a stylized hypothetical decision environment where we can randomize the number of women on a board and control for the availability of qualified candidates (Hypothesis 1b). In Studies 2A and 2B, we seek evidence that scrutiny, descriptive social norms about diversity, and goal thresholds influence the gender of group members selected for empty positions, and we experimentally manipulate social norms and scrutiny to test Hypotheses 1b and 2. Finally, in Study 3, we examine how social norms and group visibility affect the race of group members selected for empty positions, and we do this by experimentally manipulating social norms and visibility to test Hypotheses 1b and 3. Together, these studies help establish the external validity, internal validity, and generalizability of our theories.

STUDY 1: CORPORATE BOARDS

We first test our theories in the context of U.S. corporate boards. This is an important organizational setting that is economically significant, as boards control trillions of dollars. It is also highly policy relevant, as in recent years, numerous countries have passed laws about the gender composition of the corporate boards of public companies (Bainbridge & Henderson, 2014; Forbes & Milliken, 1999; Smale & Miller, 2015).

STUDY 1A: CLUSTERING OF U.S. CORPORATE BOARD COMPOSITIONS AT THE

SOCIAL NORM

In Study 1A, we analyzed the most recent available S&P 1500 corporate board composition data (from 2013) to test whether descriptive social norms influence board composition. Given the importance of scrutiny to our theoretical model (see Hypothesis 2), we first sought to establish which dimensions of corporate board diversity faced scrutiny at the time of data collection. An analysis of news articles from 2013 in the news database Lexis Nexis revealed that of 98 newspaper articles that mentioned "board diversity", 97% mentioned gender diversity, while 18% mentioned racial or ethnic diversity (the second most frequently mentioned social category). In addition, several countries in Europe have recently passed laws mandating minimum levels of gender diversity on the boards of public companies under their jurisdiction (Smale & Miller, 2015), but no such laws have been passed about other types of diversity. Given that the majority of attention regarding diversity on corporate boards focuses on gender diversity, in this study, we therefore test for (and only expect to observe) social norm effects pertaining to the gender diversity of U.S. corporate boards.

On S&P 1500 corporate boards, the average number of women was 1.36 in 2013, and this descriptive social norm received significant media coverage, with all newspaper articles in the Lexis Nexis database about board gender diversity in 2013 focusing on the average number or percentage of women on boards. We therefore expect to observe an excess of boards with exactly two women, as boards with two women just exceed the peer norm for gender diversity (Hypothesis 1a). We also predict that this excess of exactly two women per board will be more prevalent among more visible companies–those that receive more overall media attention (Hypothesis 3).

Methods

Data. Our dataset was compiled by Institutional Shareholder Services (ISS). The ISS Directors Data we analyzed contains detailed information about the boards of directors for 1,514 companies that represent the S&P Composite 1500, which is composed of three indices: the S&P 500, the S&P MidCap 400, and the S&P SmallCap 600. The S&P 1500 represents roughly 90% of the total U.S. stock market capitalization, and we also focus on the far more visible subset of companies in the S&P 500,[2] which represents roughly 90% of the total market capitalization of the S&P 1500 and 80% of the total market capitalization of the U.S. stock market (S&P Dow Jones Indices, 2015).

The ISS dataset we analyze includes information on the individual members of the boards of directors for each of the 1,514 companies in the S&P Composite 1500, including each director's name, gender, and ethnicity.[3] The dataset is updated annually, and for our primary analysis, we relied on the 2013 data, as this was the most recent data available to us as of June 5, 2015 when we first accessed the ISS database.

Additional data were collected on each company's media mentions (from Lexis Nexis), industry (from NASDAQ), year of IPO (from Bloomberg and company websites), market capitalization (from the Center for Research in Security Prices and Google Finance), and percent institutional ownership (from Bloomberg), and these data were used to perform robustness checks and investigate the moderating effect of visibility.

Analysis Strategy. To test Hypothesis 1a, we relied on a comparison of the *actual* distribution of male and female directors on corporate boards with the distribution we would

[2] A Google search for the term "S&P 500" returns 400 times as many results as a Google search for the term "S&P 1500", and a Google Scholar search for the term "S&P 500" returns 20 times as many articles as a search for the term "S&P 1500".

[3] ISS data on director gender was complete, but in 31 instances, director ethnicity was missing or blank. We manually searched Google and company websites to fill in these missing data.

expect if those directors were assigned to boards in a gender-neutral manner. We determined the expected distribution using a Monte Carlo simulation method (Rubinstein & Kroese, 2011). Specifically, we took existing 2013 S&P 1500 and S&P 500 data on directors and board seats from the ISS Directors dataset and then randomly reassigned directors to different boards, generating 10,000 simulated distributions of directors to boards. Because we *randomly* reassigned actual directors to boards in each of our simulations, these simulations produced the board composition distribution we would expect to see if gender played no role in board member selection. In other words, given the available pool of board seats and directors, our simulations told us how many women we should expect to see on each board if boards ignored gender when selecting board members.

We reassign *existing* directors in our simulations to provide a conservative test of whether there exist anomalous sorting patterns of female directors to boards.[4] In each simulation, we took as given the number of boards, the size of each board, and the number of board seats each director held based on the statistics we observed in the 2013 ISS Directors Dataset. For example, if company Alpha had nine board members in the ISS Directors Dataset, then in each simulation, company Alpha was assigned nine distinct board members. Similarly, if director Zed held two different board seats in the ISS Directors Dataset, then director Zed ended each simulation holding seats on two different corporate boards.

Running this simulation 10,000 times produced random assignments of all directors to all

[4] One common explanation for the limited number of women on corporate boards is that there are not enough qualified women to serve on boards. We thus assume the universe of people qualified to serve on boards consists only of those who actually sit on boards, so our simulations gauge whether we find anomalous sorting even if we assume no more qualified women exist to serve on boards. This extremely conservative assumption is certainly incorrect, but given that the universe of qualified women must be larger than the set who already serve on boards, finding evidence of clustering at the social norm under our assumptions would be even *more* remarkable (since relaxing this assumption would make it easier for the observed gender distribution to deviate from our simulated expected distribution).

boards that reflected the same number of directors, number of boards, and the same board sizes we observed in the ISS Directors Dataset. For each simulation result, we considered how many company boards were assigned zero female directors, one female director, two female directors, etc. We then calculated the mean of these values across all 10,000 simulations. These means told us how many companies we would expect, on average, to observe with exactly zero, one, two, and so on female directors if available board seats in the ISS dataset were randomly assigned to available directors. Our simulations also told us how rare a given assortment was, giving us bounds in the form of confidence intervals around each mean to indicate the likelihood under random assignment that we would observe a certain fraction of boards containing a specific number of women (e.g., in what fraction of 10,000 simulations had we obtained such a result).

Although this simulation strategy has been used and validated in a number of empirical papers (e.g., Dezső, Ross, & Uribe, 2016; Gino & Pierce, 2010), we also conducted placebo simulations with a characteristic other than gender to ensure that any observed deviations from our simulations on gender were not an artifact of our simulation method (see *Robustness Checks*).

Results

Summary Statistics. For companies in our dataset, the modal number of directors on a board was nine, the median number was nine, and 95% of companies had between 6 and 14 directors. Because we were interested in understanding the distribution of the absolute number of women on each board, boards with outlier numbers of seats could have exerted undue influence on our analyses. For our primary analyses, we therefore trimmed our dataset to include only companies with a total number of directors in the middle 95% of the distribution, excluding companies with outlier numbers of directors (i.e. fewer than six or more than 14) and leaving us

with 1,441 companies to analyze. However, the results of our analyses remain meaningfully unchanged in terms of magnitude and statistical significance if we repeat them without trimming these outliers (see *Online Supplement*).

The 1,441 companies in our trimmed data set included 13,440 distinct board seats and 11,185 distinct directors, as some directors held board seats on multiple company boards. In our trimmed dataset, 84% of directors held exactly one board seat; 13% held two board seats; 3% held three board seats; and less than 1% held four or five board seats. Of the 11,185 unique directors represented in our trimmed dataset, 14% (1,558) were female, and women held 15% (1,963) of the available board seats (see Table 1). Ninety-one percent (N = 10,150) of directors were Caucasian, 3.7% (N = 417) were Black, 3.0% (N = 335) were Asian, 1.7% (N = 192) were Hispanic, and 0.8% (N = 91) were classified as belonging to a different ethnic group (see Table 1). The average age of the directors in our trimmed dataset was 62.9 years with a standard deviation of 8.9 years. Fifty-eight (4.0%) of the companies had female CEOs. See Table 2 for a correlation matrix describing our data.

Insert Table 1 about here

Insert Table 2 about here

Do Boards Cluster Around the Descriptive Social Norm for Gender Diversity?

Hypothesis 1a suggests we should find an excess of boards with exactly two women (since the relevant descriptive social norm was that an average board in the S&P 1500 included 1.36 women in 2013 and an average board in the S&P 500 included 1.89 women in 2013). Based on simulations of the S&P 1500, there were 8% fewer companies with no women than would be expected ($p < 0.02$), and consistent with Hypothesis 1a, there were 12% more boards with

exactly two women than would be expected ($p < 0.01$). Boards including other frequencies of women were in line with expectations (see Figure 1, Panel A). Similarly, for the S&P 500 and consistent with Hypothesis 1a, there were 45% more companies with exactly two female board members than would be expected ($p < 0.001$). There were also 45% fewer companies with no female board members than we would expect ($p < 0.001$), and boards including other frequencies of women again arose at the rate expected (see Figure 1, Panel B). Thus, Hypothesis 1a is supported, and in light of the far higher visibility of S&P 500 companies than other companies in the S&P 1500, these patterns provide suggestive evidence in support of Hypothesis 3.

Insert Figure 1 about here

To provide further support for Hypothesis 1a, we analyzed additional historical data on corporate board composition to assess whether historical descriptive social norms also determined where clustering occurred. In years when the average number of women per board (i.e. the descriptive social norm) was below one woman, our theorizing predicts an overrepresentation of boards with exactly one woman (i.e. "tokenism" or a group including exactly one woman (Kanter, 1977)); in years when the average number of women per board was between one and two women (e.g., 1.36 women per board in 2013), our theorizing predicts an overrepresentation of boards with exactly two women. We name the phenomenon whereby a group includes exactly two women "twokenism", which is a portmanteau of the number "two" and the term "tokenism" originally used by Kanter (1977). We repeated our simulations using twelve years of historical data to see if the descriptive social norm did in fact predict where an excess of boards arose in each distribution.

We gathered additional data on the composition of S&P 1500 boards from 2002 to 2012

from the RiskMetrics Directors Legacy dataset (for the years 2002 to 2006)[5] and the ISS (RiskMetrics) Directors dataset (for the years 2007 to 2012) on August 22, 2016. For each year from 2002 to 2012, we repeated our simulation strategy to calculate how many boards would be expected to include exactly one or exactly two female directors and then compared these simulation-based expectations to the number of boards we actually observed with exactly one or exactly two female directors.

As illustrated in Figure 2, we found a statistically significant overrepresentation of boards with exactly one woman when the descriptive social norm was below one woman per board and statistically significant overrepresentation of boards with exactly two women when the descriptive social norm rose above one woman per board. In 2002 and 2003, the descriptive social norm for gender diversity–or the average number of women per board–was less than one woman, and we see statistically significant tokenism in these two years, but we do not find statistically significant twokenism in these years. From 2005 to 2013, the descriptive social norm for gender diversity exceeded one woman, and we see statistically significant twokenism in these years, but we do not find statistically significant tokenism in these years, however. In 2004, the first year that the descriptive social norm for gender diversity exceeded one woman in the S&P 1500, we still observe statistically significant tokenism and do not yet find statistically significant twokenism.

When we ran an ordinary least squares regression with robust standard errors clustered at the firm level to predict the extent of tokenism (or the overrepresentation of boards including one

[5] Data captured prior to 2002 in the RiskMetrics Directors Legacy dataset appear to have substantial variation in data quality and reliability. For example, although the dataset is meant to include information about S&P 1500 companies, and there are roughly 1500 companies in the S&P 1500, the 2001 dataset included information about 1797 companies supposedly in the S&P 1500, suggesting it was unreliable. This is why we began our analyses with data from 2002. ISS Director data is only available going back to 2007.

woman) or twokenism (or the overrepresentation of boards including two women) in each year as a function of whether the descriptive social norm for gender diversity exceeded one woman in that year, we found that the descriptive social norm exceeding one woman was a significant *negative* predictor of tokenism (β = -0.11; $p < 0.001$) and a significant *positive* predictor of twokenism (β = 0.12; p = 0.002). This provides further support for Hypothesis 1a and our theorizing that descriptive social norms help determine salient thresholds for diversity.

Insert Figure 2 about here

Are More Visible Companies More Likely to Exhibit Twokenism? To test Hypothesis 3 in this context, we examined whether companies that receive more media attention were more likely to include exactly two women on their boards. We used media attention as a proxy for visibility to align with past research on organizational visibility (Brammer & Millington, 2006; Chiu & Sharfman, 2011; King, 2008; McDonnell & King, 2013). We searched Lexis Nexis for all media mentions (including newspapers, web-based publications, magazines, etc.) of each of the companies in the S&P 1500 in 2012 (mean media mentions of a company = 307; S.D. = 441). We gathered 2012 data on media attention so we could examine whether past media attention predicted future (2013) twokenism. We then analyzed whether media attention in 2012 predicted whether companies would include exactly two women on their boards in 2013.

We ordered the companies in our dataset by the number of media mentions each company received in 2012 and created deciles (i.e. ten bins of 144 companies each) based on this ordering. Thus, the first decile contained the companies most frequently mentioned in the media in 2012, while the last decile contained the companies least frequently mentioned in the media in 2012. After segmenting the companies in our dataset by the amount of media attention they were

subjected to in 2012, we repeated our basic simulation strategy but limited each simulation to include only the companies in a given decile. This allowed us to determine how many companies we would expect to see with exactly two women on their boards in 2013 in each of the deciles. We ran 1,000 simulations for each decile, generating a new expected number of companies with exactly two female directors each time. Thus, for each decile, we generated an *expected* number of companies with exactly two women on their boards based on our simulations, and we could compare this with the *actual* number of companies including exactly two women on their boards in our 2013 board data.

The results of our simulations for the different media attention deciles are depicted in Figure 3. To test the hypothesis that the likelihood of having exactly two women on a company's board increases for more visible companies, we ran an ordinary least squared (OLS) regression with robust standard errors. We used the logarithm of the average number of media mentions in a given decile to predict the absolute difference between the observed and expected number of companies with exactly two women on their boards in each decile. The logarithm of media mentions of the decile was a significant predictor of the absolute difference between observed and expected boards with exactly two female directors ($\beta = 6.12$, $p = 0.014$). The positive coefficient of log media mentions indicates that deciles containing more visible companies were more likely to display twokenism, supporting Hypothesis 3.[6]

Insert Figure 3 about here

---

[6] See the *Online Supplement* for additional specifications of this regression to test the robustness of this finding and for a table reporting detailed regression results. We used as predictors either the logarithm of the average number of media mentions or the decile rank, and we used as outcomes either the absolute overrepresentation of boards with exactly two women or the percent overrepresentation of boards with exactly two women. All yielded findings that were statistically significant and meaningfully unchanged.

Robustness Checks. To further validate our simulation strategy and ensure our results were not an artifact of the way we constructed an expected distribution of the number of boards including varying numbers of female directors, we conducted a series of placebo simulations (Gino & Pierce, 2010). Specifically, in these placebo simulations, we produced expected distributions of the number of boards that would include varying numbers of directors with another characteristic (i.e., not gender) that should not show goal-related clustering effects because of a lack of scrutiny on that characteristic (e.g., board members whose ages ended with an arbitrary number). We found no significant differences between the expected numbers of boards and the actual numbers of boards in any of our placebo simulations, suggesting that the large deviations we see in our simulations studying gender were not an artifact of the way we constructed our baseline expectations or null distributions (see the *Online Supplement* for complete details about our placebo simulations).

In addition to conducting placebo simulations to ensure the robustness of our simulation methodology, we conducted numerous additional robustness checks to ensure our results were not driven by outliers or by a small subset of boards by repeating our baseline simulations with different cuts of our data. First, we checked that our findings were robust to board size. To do this, we used our standard simulation strategy but limited the data to boards of size 6 or fewer, 7, 8, 9, 10, 11, 12, or 13 or more. The underrepresentation of companies with no women on their boards and the overrepresentation of companies with exactly two women on their boards is robust across all board sizes tested (see Table 3), although the clustering at the social norm of two is largely driven by companies with larger boards, and future research exploring the reasons for this could yield interesting insights.

Insert Table 3 about here

Our results are also robust across industries, and they hold regardless of the gender of a company's CEO (see *Online Supplement*). We also examined whether the length of time the company has been public affects the likelihood that the company has exactly two women on its board. In general, our results appear to be robust regardless of when a company went public (see *Online Supplement*). Finally, when we examine how our results relate to the market capitalization of a company, we find that twokenism is more prevalent among companies with higher market capitalization (see *Online Supplement*), which are also the most frequently mentioned by the media (the correlation between the logarithm of a company's market capitalization and the logarithm of its number of media mentions in 2013 = 0.59; $p < 0.001$).

STUDY 1B: THRESHOLD EFFECTS IN BOARD MEMBER SELECTION AT THE SOCIAL NORM

In Study 1B, we analyzed the gender of new board members added to company boards over time for evidence consistent with our theories. We predicted that boards would be discontinuously less likely to add additional women once they had met the relevant descriptive social norm for gender diversity (Hypothesis 1b). Given that the descriptive social norm for gender diversity in the S&P 1500 first surpassed one woman in 2004, we examine all board member additions from 2004 to 2013 to test whether boards during this time period were discontinuously less likely to add additional female directors once they already included two women on their boards.

Method

Data. For these analyses, we use a subset of the data described in Study 1a. Specifically, we use the ISS Directors dataset describing board composition from 2007 to 2013 and the

RiskMetrics Directors Legacy dataset describing board composition from 2004 to 2006 to examine the 9,989 board member additions in the S&P 1500 from 2004 to 2013.

Analysis Strategy. Using data on all board member additions from 2004 to 2013, we estimated an ordinary least squares regression with robust standard errors to predict whether each newly added board member was female.[7] We included as predictors both the number of women currently on a board (to control for the possibility that boards have either increasing or decreasing marginal value for female directors) as well as an indicator for whether the board included at least two women (our primary predictor of a discontinuity in a groups' desire to add more women after exceeding the social norm for gender diversity), and we clustered standard errors by firm. We report these regressions with and without fixed effects for board size, fixed effects for industry, fixed effects for stock market index, and a continuous control for a company's market capitalization.

Results

Summary Statistics. Of the 9,989 board additions from 2004 to 2013, 16.5% (1,649) were additions of female directors. The 9,989 board member additions from 2004 to 2013 represent 8,328 distinct directors (i.e. some directors were added to multiple boards during this time period), and 16.2% (1,347) of the distinct directors were female. On average, boards in this dataset added 5.25 directors during this nine-year span.

Do Boards Add Fewer Women Once They Have Reached the Descriptive Social Norm? As shown in Table 4, Model 1, for the S&P 1500, the coefficient on our primary predictor of whether a board added a female director–an indicator for whether the board already included at

[7] We rely on a linear model because it yields more interpretable coefficients than a logit specification, and this method also allows us to correct for heteroskedasticity in the standard errors (Angrist & Pischke, 2008; see Brands & Fernandez-Mateo, 2016 for a similar procedure). However, logistic regressions yield similar results and are reported in the *Online Supplement*.

least two women–was negative and significant in our primary regression specification (β = -0.039; p = 0.017). As shown in Table 4, Model 3, for the S&P 500 (roughly the 500 most visible and valuable companies in the S&P 1500), the coefficient on the indicator variable was negative and even more highly significant (β = -0.092; $p < 0.001$). This suggests that companies are less likely to add additional women to their boards once their boards have met the social norm for gender diversity by including two women, providing support for Hypothesis 1b. The larger effect size in the (highly visible) S&P 500 also provides some suggestive support for Hypothesis 3. Adding in fixed effects for board size, fixed effects for industry, fixed effects for stock market index, and a continuous control for market capitalization (see Table 4, Models 2 and 4), we still find that our predictor of a discontinuity is significant in the S&P 1500 (β = -0.034; p = 0.037) and in the S&P 500 (β = -0.090; $p < 0.001$).

Insert Table 4 about here

Do More Visible Companies Show Larger Discontinuities at the Descriptive Social Norm? To test Hypothesis 3 in Study 1B, we examined whether there was an interaction between media attention and our primary predictor of whether a board added a female director–an indicator for whether the board already included at least two women. We again searched Lexis Nexis for all media mentions of each of the companies in the S&P 1500, and we gathered additional data to look at media mentions for each year starting in 2004 to see if media attention in year $t - 1$ predicted whether a newly added board member in year t was female. For our analyses, we used the centered logarithm of media mentions rather than the raw number of media mentions because the distribution of media mentions is highly skewed (skewness = 2.57; skewness test for normality $p < 0.0001$; kurtosis = 10.37; kurtosis test for normality $p < 0.0001$).

Our results are depicted in Table 5. As predicted, we find a significant negative interaction between the centered logarithm of the number of media mentions in year *t – 1* and having two or more women on a board in predicting whether a newly added board member in year *t* was female (β = -0.021; p = 0.042; Model 1). Adding in fixed effects for board size, fixed effects for industry, fixed effects for stock market index, and a continuous control for market capitalization, we still find a significant negative interaction between the centered logarithm of media mentions and having two or more women (β = -0.021; p = 0.041; Model 2). These results suggest that more visible companies show larger discontinuities in board member additions at the descriptive social norm of two women per board.

Insert Table 5 about here

STUDY 1C: ONLINE EXPERIMENT REPLICATING THRESHOLD EFFECTS

In Study 1C, we sought to replicate our findings regarding threshold effects from Study 1B in an online experiment that allowed us to randomly assign the number of women in a group and control for the availability of qualified candidates. Specifically, we investigated whether individuals in a controlled setting are less likely to add women to a corporate board when the board has met or surpassed the social norm for gender diversity by including two or more women.

Method

Participants. Four hundred and seventy-nine U.S. participants were recruited through Amazon's Mechanical Turk to participate in a short online research study (55% male; 77% Caucasian). These participants were paid $0.25 for completing a survey they were told would take approximately 5 minutes of their time. Sample size was determined a priori, data analysis

was conducted only once all data were collected, and we do not exclude any data.

Procedures. In a pilot study (see the *Online Supplement* for details), we first established that our study population was indeed aware that two is the average number of women on U.S. corporate boards (i.e. two women is the descriptive social norm for gender diversity).

After establishing an awareness of descriptive social norms in a pilot, we ran our primary study. In this study, participants were asked to imagine they had been tasked with helping a company select a new member for its board of directors. They were then exposed to a list of ten names and told the current board consisted of the individuals on that list. Participants were randomly assigned to one of four experimental conditions where zero, one, two, or three of the names of board members were female.

Study participants were next presented with three hypothetical candidates for an opening on the board in question and asked to choose one to add to the board. The candidates were all described as qualified, but one was a CEO, one was a current board member at another company, and one was a consultant with expertise in the industry. We randomly varied which candidate had a female name (Jill Davis) and which candidates had male names (Matthew Anderson and Todd Miller), and we randomly varied which name was associated with each qualification.[8] We presented three candidates for the available board seat rather than one male and one female to reduce suspicion that our study was about gender following Castilla and Benard (2010). Our dependent variable of interest was what fraction of participants in each condition would choose a female candidate.

Finally, participants completed demographic questions and a manipulation check

[8] Participants were most likely to choose the candidate who was a CEO ($p < 0.001$), regardless of gender. However, because we randomly assigned qualifications to the candidates, we do not need to control for candidate qualification in order for our tests to provide unbiased estimates of the causal effects of our manipulations. In addition, we did not find any significant interactions between gender and candidate qualifications.

question, which asked them to recall how many men and how many women were present on the corporate board they had seen at the beginning of the survey. Study materials and a correlation matrix of all variables collected in this study are available in the *Online Supplement*.

Results

First, our manipulation check confirmed our manipulation was successful: participants recalled significantly more women on boards that included three women than two ($p < 0.001$), two women than one ($p < 0.001$), and one woman than zero women ($p = 0.015$).

Second, a Chi-square test of independence showed a marginally significant relationship between the number of women on the board and whether the participant chose the female candidate ($\chi^2(3, N = 479) = 7.51$, $p = 0.057$). Consistent with Hypothesis 1b and replicating our results from Study 1B, participants were significantly less likely to choose the female candidate and increase the gender diversity of the board once the board included at least two women. Participants shown a corporate board with exactly two female members were significantly less likely to choose the female candidate for the open seat ($M = 36.0\%$, $SD = 0.482$) than were participants who were shown a corporate board with one female member ($M = 50.4\%$, $SD = 0.502$; $t(239) = 2.27$, $p = 0.024$; see Figure 4).[9] We then ran an ordinary least squares regression (OLS) with robust standard errors to predict the likelihood a participant chose the female candidate, replicating our empirical analyses of board member additions from S&P 1500 and S&P 500 data from Study 1B. We again included the number of women currently on the board as a control variable in addition to an indicator variable for whether the board included at least two women as a predictor of a discontinuity. The coefficient on the indicator variable was negative and marginally significant ($\beta = -0.187$, $p = 0.062$; see Table 6), suggesting that participants in

[9] We find a main effect of participant gender such that female participants are significantly more likely to select the female candidate ($p = 0.019$), but we find no significant interaction between participant gender and decisions.

our experiment also were discontinuously less likely to increase the gender diversity of the board once the board had at least two women and providing additional support for Hypothesis 1b.

Insert Figure 4 about here

Insert Table 6 about here

DISCUSSION OF STUDY 1

Study 1A shows that U.S. corporate boards are disproportionately likely to include exactly the number of women needed to minimally exceed the descriptive social norm for female representation in peer groups. This evidence is consistent with Hypothesis 1a, which proposes that the composition of groups facing scrutiny along a diversity dimension will cluster around the descriptive social norm for that type of diversity. Further, historical analyses show that descriptive social norms predicted the shift from tokenism (an overabundance of boards with exactly one female director) to twokenism (an overabundance of boards with exactly two female directors), providing additional support for Hypothesis 1a that the clustering we detect is driven by the descriptive social norm for gender diversity.

Study 1B provides support for Hypothesis 1b, which states that groups facing scrutiny along a diversity dimension will be less likely to add members of the relevant underrepresented group once they have reached the descriptive social norm for diversity. We find that U.S. corporate boards are discontinuously less likely to add additional women once they have reached the descriptive social norm for diversity by including two female directors. In Study 1C, we replicate this finding in a stylized experiment where we randomly assign the number of women to a hypothetical corporate board and control for the availability of qualified candidates. While Study 1C lacks the realism of Studies 1A and 1B, it confirms our hypothesis in an environment

where we can randomly assign board composition, providing convergent evidence that there exists a causal relationship between board composition and the gender of new board members.

Consistent with Hypothesis 2, which predicts scrutiny is a necessary condition for social norms to influence diversity, we do not see evidence of clustering at the social norm when we look at board members' race or ethnicity in supplemental analyses.[10] There is far less scrutiny of corporate boards' racial diversity compared with the scrutiny boards face regarding gender diversity (i.e., only 18% of news articles about board diversity in 2013 discussed racial diversity while 97% discussed gender diversity, and no laws have been passed establishing racial quotas on corporate boards in any country), so corporate boards may have fewer impression management motives regarding the recruitment of racial or ethnic minorities compared to women.

Finally, consistent with Hypothesis 3, we find evidence that more visible companies (as measured by media coverage in the previous year) are more likely to include exactly two women on their boards, consistent with our theory that the clustering we detect at the social norm is driven in part by impression management concerns. In Study 1B, we also find that more visible companies show larger discontinuities at the descriptive social norm of two women per board when adding additional female board members.

Past research suggests these findings are worrisome from a policy perspective. Research on the benefits of gender diversity on corporate boards suggests that at least three female directors are needed before boards experience tangible benefits from gender diversity (Konrad, Kramer, & Erkut, 2008; Torchia, Calabrò, & Huse, 2011). By stopping at two women, boards may be missing out on key benefits that can ensue from greater gender diversity (cf. Adams &

[10]A more detailed discussion of simulation analyses regarding director race and ethnicity can be found in the *Online Supplement*.

Ferreira, 2009; Matsa & Miller, 2013). Further, our results suggest that the push for gender parity on boards may not generate results for a long time. In Study 1A, we depict the evolution of descriptive social norms regarding gender diversity on corporate boards over a twelve-year span, and these results suggest that descriptive social norms change quite slowly over time.

In spite of the compelling evidence provided by our empirical analyses of archival board composition data supporting our theorizing and hypotheses, Studies 1A and 1B are ultimately only correlational studies and thus have limitations. We cannot completely rule out concerns about reverse causality or other confounds such as firm performance. In addition, because we do not observe board member selection decisions directly, we can only explore the mechanisms responsible for the effects we have documented indirectly. There are many factors at play that affect who is added to corporate boards (e.g., legal constraints can prevent people from serving on multiple boards; bias and stereotyping may affect board member selection), and we focus only on the roles played by descriptive social norms, scrutiny, and visibility. We also unfortunately cannot disentangle the specific motives of individual companies.

In order to provide more confidence in our results, in Study 1C, we replicated threshold effects at the descriptive social norm in an experimental setting where we could randomize the number of women in a group. This gives us greater confidence that the results found in Study 1B are not driven by endogeneity or the fact that there are not enough qualified women for director positions. However, we acknowledge that Study 1C is a stylized experiment that does not accurately represent corporate board decision-making processes. First, our experiment is conducted at the level of the individual, while boards are groups. Second, board members have much more experience and many more constraints they must attend to, while we use a relatively uninformed sample and intentionally stripped away many of the complications of the board

member selection process for simplicity.

In spite of these limitations, these studies collectively provide empirical evidence that group composition and group diversity decisions can be affected by threshold effects at the descriptive social norm. In our following studies, we provide additional experimental evidence directly testing our theoretical model to examine the influences of descriptive social norms, scrutiny, and visibility on group diversity decisions.

STUDY 2

In Study 2, we sought to test our theoretical model more directly by manipulating–rather than measuring–the descriptive social norms and scrutiny associated with the inclusion of females in a group. In addition, we sought to explore these phenomena in a new setting to establish their generalizability to groups besides corporate boards.

STUDY 2A: GROUP DIVERSITY, SOCIAL NORMS AND SCRUTINY

In Study 2A, we tested whether manipulating descriptive social norms and scrutiny affects decisions about whether to add a female candidate to a majority-male group with a sample of participants with work experience. Specifically, we investigated whether, as predicted in Hypothesis 2, individuals strive to meet descriptive social norms for diversity when under threat of possible scrutiny but not in cases where scrutiny is absent.

Method

Participants. Five hundred and fifty-six Master of Business Administration (MBA) students completed this study. This represented the entire incoming class at a U.S. business school. 57% of the participants were male, 25% had previous managerial experience before starting their MBA, and participants' average age was 27.7 years. Sample size was determined a priori, data analysis was conducted only once all data were collected, we do not exclude any

data, and we report all measures and manipulations.

Procedures. Participants were asked to imagine their company had given them the task of assembling a seven-person panel for submission to an industry conference. They were told six of the seven panelists had already been determined, and they were responsible for selecting the final panelist. All participants saw an image of two women and four men representing the six predetermined panelists. Participants were randomly assigned to one of four experimental conditions (*surpassed social norm* or *unmet social norm* x *unscrutinized* or *scrutinized*) described below.

Participants saw images of five seven-person panels representing other panel submissions to the conference. Participants randomly assigned to the *surpassed social norm* condition saw that four of these other panel submissions had one woman each while one panel submission had no women on it (i.e. the average number of women on other panels was 0.8); participants randomly assigned to the *unmet social norm* condition saw that four of these panel submissions had three women each while one panel submission had two women on it (i.e. the average number of women on other panels was 2.8). Therefore, in the *surpassed social norm* condition, the participant's current panel (which included two women) already exceeded the descriptive social norm for gender diversity (0.8 women); in the *unmet social norm* condition, the participant's current panel (which included two women) was below the descriptive social norm for gender diversity (2.8 women).

Participants were told panels were generally accepted based on speaker quality and years of industry experience of the panelists. Participants randomly assigned to the *unscrutinized* condition were told the review process was "blind": the names and photos of the panelists would not be submitted for evaluation (i.e. it would be impossible for the panels to be scrutinized with

regards to gender composition). Participants randomly assigned to the *scrutinized* condition saw no such statement. Past research suggests that impression management concerns often arise when people simply know they are being evaluated (Leary & Kowalski, 1990), suggesting that when the evaluation process is not blind, scrutiny can be expected to affect decisions.[11]

Participants were then shown two potential candidates–Candidate A and Candidate B–for the final panelist. One image depicted a female candidate who had 10 years of industry experience and a speaker rating of 4.6; the other image depicted a male candidate who had 12 years of industry experience and a speaker rating of 4.8. Which candidate was presented first as Candidate A (versus second as Candidate B) was randomized. Participants then rated their preference for the two candidates on a scale from 1 to 7 where 1 was labeled as "Strongly prefer Candidate A" and 7 was labeled as "Strongly prefer Candidate B." Study materials and a correlation matrix of all variables collected in this study are available in the *Online Supplement*.

Results

Consistent with Hypothesis 1b and as illustrated in Figure 5A, participants in the *scrutinized* condition had a significantly stronger preference for the female candidate in the *unmet social norm* condition than in the *surpassed social norm* condition ($t(277) = 2.24$; $p = 0.026$). In other words, participants whose diversity decisions could be scrutinized found it much more desirable to add a female candidate to a group when the group had not yet met the social norm for gender diversity compared to when the group had surpassed the social norm. However, consistent with Hypothesis 2, there were no differences in the preferences expressed for the

[11] In a separate pilot study, we asked participants to rate how much they agreed or disagreed on a 7-point scale with the statements, "My decision is under scrutiny with regards to the gender diversity of the panel" and "The reviewer will pay attention to the gender diversity of the panel when deciding which panels to accept". Participants in the *scrutinized* condition reported significantly higher scrutiny on gender diversity than participants in the *unscrutinized* condition ($M_{scrutinized} = 3.67$, $SD_{scrutinized} = 1.84$; $M_{unscrutinized} = 2.66$, $SD_{unscrutinized} = 1.92$, $t(150) = 3.34$, $p = 0.001$).

female candidate between the *unmet social norm* and the *surpassed social norm* conditions when diversity decisions were not under scrutiny ($t(275) = 0.216$; $p = 0.829$).

Insert Figure 5 about here

Next, we checked if there was a significant interaction between surpassed social norms and the presence of scrutiny. We estimated a linear regression to predict the preference for the female candidate with indicators for our *scrutinized* condition, our *unmet social norm* condition, and the interaction between these two conditions (see Table 7, Model 1). The interaction term was positive but did not reach standard levels of statistical significance ($p = 0.140$).[12] To strengthen our statistical power to detect an interaction, we conducted a follow-up study with incentivized decisions (note that we could not increase the sample size in this study because it already included every incoming MBA student at our selected university, and we were not able to incentivize the decisions of MBA students).

Insert Table 7 about here

STUDY 2B: REPLICATING STUDY 2A WITH INCENTIVES

In Study 2B, we sought to replicate our results from Study 2A but with real monetary stakes that would increase our statistical power to detect an interaction between the presence of scrutiny and a surpassed social norm for diversity. Again, we experimentally manipulated scrutiny and descriptive social norms to test for a causal relationship between these variables and the demographic characteristics of a newly-selected group member.

Method

[12] We found a significant main effect of gender such that women had significantly higher preferences for the female candidate compared to men ($p = 0.022$), but there was no significant interaction.

Participants. Two hundred U.S. participants (51.5% male) were recruited through Amazon's Mechanical Turk and paid $0.15 to participate in a short online research study. Sample size was determined a priori, data analysis was conducted only once all data were collected, we do not exclude any data, and we report all measures and manipulations.

Procedures. As in Study 2A, participants were asked to imagine their company had given them the task of assembling a seven-person panel for submission to an industry conference, that six of the seven panelists had already been determined (two women and four men), and that they were responsible for selecting the final panelist. Again, they were randomly assigned to one of four experimental conditions.

In this study, we simplified the way the descriptive social norm was manipulated. Participants randomly assigned to the *surpassed social norm* condition were told competitive intelligence suggested the other panel submissions would have 1.25 women on average; participants randomly assigned to the *unmet social norm* condition were told the other panel submissions would have 2.75 women on average.

Participants were then told a reviewer would evaluate all panel submissions and choose to "accept" 75% of them. If their panel submission was accepted, participants would receive a bonus payment. All participants were initially allocated a $0.25 bonus, but participants had to "pay" to select the last panelist, and this cost was deducted from their promised bonus. Participants randomly assigned to the *unscrutinized* condition were told the review process was "blind": the names and photos of the panelists would not be submitted for evaluation (i.e. it would be impossible for the panels to be scrutinized with regards to gender composition). Participants randomly assigned to the *scrutinized* condition saw no such statement.

Participants were then offered the choice among three candidates for their final panelist.

One image depicted a female candidate who had 10 years of industry experience, a speaker rating of 4.6, and cost $0.15 to select. The other images depicted male candidates who had similar qualifications (11 or 12 years of industry experience; speaker ratings of 4.5 or 4.8) and cost $0.10 and $0.11 to select. Our outcome of interest was what fraction of participants in each condition selected the female candidate. We made the female candidate slightly more expensive to reflect research suggesting that women are more expensive to recruit and/or hire in contexts where diversity is lacking (e.g., on corporate boards and other contexts where less than 50% of the workforce is female, see Leslie, Manchester, & Dahm, 2016). Finally, participants reported their gender identity and whether they had ever attended or organized a conference in the past 10 years. Study materials and a correlation matrix of all variables collected in this study are available in the *Online Supplement*.

Results

Consistent with Hypothesis 1b, participants in the *scrutinized* condition were significantly more likely to select the female candidate in the *unmet social norm* condition than in the *surpassed social norm* condition ($z = 2.941$; $p = 0.0033$; see Figure 5B). Consistent with Hypothesis 2, there were no such differences in the likelihood of selecting the female candidate in the *unmet social norm* and the *surpassed social norm* conditions when there was no scrutiny ($z = 0.242$; $p = 0.808$).

To test for an interaction between the presence of scrutiny and unmet social norms, we estimated an ordinary least squares regression with robust standard errors to predict the choice of a female candidate with indicators for our *scrutinized* condition, our *unmet social norm* condition, and the interaction between these two conditions (see Table 7, Model 2). We found that the interaction term was positive and statistically significant ($\beta = 0.270$; $p = 0.028$). This

further supports Hypothesis 2 that predicts when shaping the composition of groups, decision makers will only conform to the social norm for diversity when they are under scrutiny.

DISCUSSION OF STUDY 2

Studies 2A and 2B directly manipulate scrutiny and descriptive social norms to provide direct tests of Hypotheses 1b and 2 and show that decision makers responsible for shaping group composition strive to increase group diversity when the group in question has not yet met the social norm for diversity on a scrutinized dimension (gender in the case of these studies). However, motivation to further increase diversity is reduced once the social norm has been met, and social norms do not exert this influence when scrutiny is not present.

STUDY 3: THE MODERATING EFFECT OF VISIBILITY

In Study 3, we manipulated descriptive social norms and a group's visibility to investigate whether the influence of descriptive social norms on decisions about group diversity is moderated by a group's visibility, and we also extend our study of group diversity to explore a social category besides gender.

Method

Participants. Six hundred and three U.S. participants (52.9% male; 80.4% Caucasian) were recruited through Amazon's Mechanical Turk to participate in a short online research study in exchange for $0.30. Sample size was determined a priori, data analysis was conducted only once all data were collected, we do not exclude any data, and we report all measures and manipulations.

Procedures. Participants were told to imagine they were the manager of a team of five people and were hiring a sixth team member. All participants saw an image of one black man and four white men representing their current team. They were also told their HR department

cared about the racial diversity of teams and the HR department could review team compositions and choose to punish teams deemed to have inadequate racial diversity, creating scrutiny on the dimension of racial diversity in all conditions. Participants were then randomly assigned to one of four experimental conditions.

Participants randomly assigned to the *surpassed social norm* condition were told that other teams of their size included an average of 0.25 black people. Participants randomly assigned to the *unmet social norm* condition were told that other teams of their size included an average of 1.75 black people.

To manipulate visibility, participants were either randomly assigned to learn either: (1) their team was "not very important" in the company so there was a low probability that the HR department would review the composition of their team (the *low visibility* condition); or (2) their team was "very important" in the company so there was a high probability that the HR department would review the composition of their team (the *high visibility* condition).[13]

Participants were then offered the choice of two candidates for their new team member. One image depicted a black male candidate who would come with a bonus of $0.03 to participants if they chose him; the other image depicted a white male candidate who would come with a bonus of $0.10 to participants if they chose him. We incentivized participants to choose the white man in order to overcome social desirability concerns and place some cost on increasing diversity. Participants were told they would keep the bonus associated with the candidate they chose unless the HR department reviewed their team and chose to penalize their

[13] In a separate pilot study, we asked participants to rate how much they agreed or disagreed on a 7-point scale with the statements, "My team is visible in the company" and "My team receives a lot of attention in the company". Participants in the *high visibility* condition reported significantly higher scores on these items than participants in the *low visibility* condition ($M_{high\_visibility} = 6.39$, $SD_{high\_visibility} = 0.97$; $M_{low\_visibility} = 2.25$, $SD_{low\_visibility} = 1.49$; $t(147) = 20.04$; $p < 0.0001$).

team for a lack of racial diversity.

Finally, participants reported their racial and gender identities. Study materials and a correlation matrix of all variables collected in this study are available in the *Online Supplement*.

Results and Discussion

Consistent with Hypothesis 1b and all previous studies, participants were significantly more likely to select the black candidate in the *unmet social norm* condition than in the *surpassed social norm* condition ($z = 4.279$; $p < 0.0001$; see Figure 6). In other words, decision makers added the black candidate to their group at a lower rate once their group had surpassed the descriptive social norm for racial diversity. In addition, there was a significant main effect of visibility, such that participants were significantly more likely to select the black candidate when their team was highly visible than when it was not ($z = 9.247$; $p < 0.0001$).

Insert Figure 6 about here

Insert Table 8 about here

To test Hypothesis 3 that visibility moderates the effect of descriptive social norms, we tested for an interaction between visibility and social norms. To do this, we estimated an ordinary least squares regression with robust standard errors to predict the choice of the black candidate with indicators for our *high visibility* condition, our *unmet social norm* condition, and the interaction between these two conditions (see Table 8). Consistent with Hypothesis 3, we found the interaction term between visibility and norms was positive and statistically significant ($\beta = 0.151$; $p = 0.043$).

Overall, Study 3 conceptually replicates our previous studies, extends our findings to underrepresented groups besides women, and shows the moderating effect of visibility on

decisions about group diversity.

GENERAL DISCUSSION

Across four experiments and one field study, we offer convergent evidence that those who shape the diversity of groups attend to and seek to conform to the descriptive social norms for diversity set by peer groups when under scrutiny. In Study 1, we showed that U.S. corporate boards are disproportionately likely to include exactly two women (the descriptive social norm), and they appear to lose motivation to add additional women once they have matched the descriptive social norm by including two female directors. We also found that these effects are more pronounced among more visible companies, consistent with our theory that these effects are driven in part by scrutiny and impression management motives. In addition, we did not find any clustering when we analyzed data on the race/ethnicity of board members in our field data, consistent with our theory that scrutiny is required to produce clustering at the descriptive social norm.[14] In Studies 2 and 3, we directly manipulated descriptive social norms, scrutiny, and visibility to show that each of these influences group diversity decisions as our theory predicts in groups besides corporate boards and when we examine social categories besides gender.

Theoretical and Practical Implications

Our theory and findings help us understand how decision makers with the power to shape group composition respond to the threat of negative scrutiny surrounding diversity. Individuals responsible for group compositions look to descriptive social norms, matching the levels of diversity found in peer groups at an unusually high rate. This behavior leads to homogeneous levels of diversity across groups, providing another contributing explanation for the persistent underrepresentation of women and racial minorities in many organizational contexts. Our work

[14] As discussed in Study 1A, an analysis of media attention to board diversity showed that 97% of such articles discuss gender diversity, while just 18% even mention racial or ethnic diversity.

also helps provide a fuller understanding of diversity-related hiring decisions, suggesting when women and racial minorities will be particularly attractive candidates for inclusion in groups and when groups can be expected to lessen their efforts to increase diversity.

Our findings suggest new avenues for policy makers seeking to increase diversity. Rather than simply targeting bias and stereotyping among those making hiring decisions (e.g., through diversity training) or seeking to shape underrepresented candidates' preferences and skill sets (e.g., by training women to negotiate), more interventions may be needed to change the perceived norms around diversity. Groups appear to cluster at the descriptive social norm for diversity because it is an adaptive impression management strategy: by clustering at the social norm, they can escape negative scrutiny regarding their diversity levels. But the fact that groups can escape negative scrutiny once they reach the descriptive social norm for diversity implies that those scrutinizing these groups (e.g., shareholders, the media, etc.) may be too easily satisfied. Shifting the standards of those who scrutinize diversity as well as those of the decision makers capable of shaping group diversity from focusing on descriptive social norms in peer groups to instead achieving more ambitious norms (e.g., matching the levels of diversity in the general population) may be a promising new avenue for increasing the diversity of highly visible, scrutinized groups. If powerful institutions or individuals endorse new norms regarding gender and racial representation, perhaps this could lead to changes in the norms that influence group composition decisions (Paluck & Shepherd, 2012). For instance, decisions by the Supreme Court have been shown to change attitudes and perceptions of norms in the realm of gay rights (Tankard & Paluck, 2017).

Our work also points to scrutiny as a lever for change. Scrutiny can come from a variety of sources, but some sources may be more influential than others (Oliver, 1991). Applying

greater scrutiny to group diversity should lead groups to increase their diversity. One extreme form of scrutiny when it comes to diversity is to enforce legal penalties on public companies for a failure to diversify. However, even when policy makers have established laws mandating minimum levels of gender diversity on the corporate boards of public companies, some companies have elected to become private rather than comply with the laws (Miller, 2014). Forced compliance therefore comes with the risk of creating at least some reactance (Dobbin, Schrage, & Kalev, 2015). An alternative to mandated diversity may be to shower positive attention on groups that reach high levels of diversity. Treating diversity as an ideal may help reshape perceptions of the relevant norm, leading injunctive norms (or norms about ideals) to overshadow descriptive social norms.

Limitations and Future Research

One paradox suggested by our theorizing and empirics surrounds changing descriptive norms: U.S. corporate boards shifted from clustering at one woman to clustering at two women (albeit slowly) over the last twenty years in spite of the fact that our theorizing about diversity thresholds would predict a stagnation of board diversity at the one-woman threshold. A noteworthy fact, however, is that this shift in clustering followed the passage of Norway's "Women on Boards" act in 2003. This legislation required public and state-owned companies in Norway to include at least 40% women and may have made the topic of gender diversity on corporate boards in the U.S. more salient at that time, providing increased scrutiny of boards with few women and making the need for gender diversity more salient, driving the shift to twokenism from tokenism. Future research exploring how descriptive social norms can be shifted in the context of diversity would be extremely valuable.

Another puzzling question raised by our findings is whether more diverse groups may

actually discriminate more than less diverse groups. We cannot evaluate whether any specific group or organization is actively "managing" diversity for impression management reasons. However, overall, we do see a pattern suggesting this is the case and suggesting that–contrary to the expectation that more diverse groups will attract more women and racial minority candidates (Avery, 2003; Avery & McKay, 2006)–such groups are less likely to select women and racial minorities than others after reaching the descriptive social norm for diversity. It would be valuable for future research to examine when and how social norms around diversity can hurt rather than help women and minorities.

Although our field and experimental studies provide convergent evidence in support of our theory and hypotheses, in our experiments, we only examine the judgments and decisions of individuals, while group member selection processes are varied and complex and often involve many decision makers. Extensive past research has shown that studies of individual decisions and insights about individual psychology can further our understanding of group and organizational outcomes (Greve, 2008; Highhouse, Brooks, & Gregarus, 2009; Simon & Houghton, 2003; Staw, 1991). However, there are unquestionably limitations in our approach.

We only test our theorizing in a single field setting (albeit in an economically and organizationally important one). Future research examining how these phenomena play out in other important organizational contexts would undoubtedly be useful. Our experiments may also be susceptible to demand effects, which could limit their external validity. In addition, in our field setting and in our experiments, the groups we examine are relatively small in size (i.e. less than 20 members). Additional research in exploring how group size moderates the effects of descriptive social norms and scrutiny could be informative. For example, in larger groups, the behavior of peer groups could feel less relevant as the size of the group might create a greater

sense of its uniqueness, reducing pressure to conform to descriptive social norms. Alternatively, larger groups may feel more scrutinized because of their size, leading them to react more dramatically to descriptive social norms.

Finally, more research into the psychological mechanisms that lead descriptive social norms and scrutiny to produce the group diversity threshold effects we document could be illuminating. Past research suggests that norms may be particularly relevant in the context of group diversity decisions because of ambiguity about how much diversity is enough and the fear of being singled out from peers (Ahmadjian & Robinson, 2001; Festinger, 1954; Sherif, 1936; Zavyalova et al., 2012). Future research isolating the specific mechanisms through which descriptive social norms exert their influence would be valuable and could help identify potent interventions for changing salient norms. Future research testing new interventions to reduce the reliance on descriptive social norms and make other norms more salient would also be extremely valuable.

Conclusion

Our work highlights the important roles that descriptive social norms, goal setting, scrutiny, and visibility play in shaping decisions about group diversity while answering questions about how individuals assess whether a group is diverse and how groups respond to scrutiny around their diversity levels. We find empirical evidence that descriptive social norms and threshold effects lead to an overabundance of groups with exactly the same level of diversity in an important organizational context, providing evidence of a previously unexplored phenomenon that may contribute to the underrepresentation of women and minorities in many organizational groups. By shedding light on novel factors that influence group diversity decisions, we illuminate potential new avenues for increasing the diversity of groups.

REFERENCES

Adams, R. B., & Ferreira, D. (2009). Women in the boardroom and their impact on governance and performance. ***Journal of Financial Economics***, *94*(2), 291–309.

Aguilera, R. V., Rupp, D. E., Williams, C. A., & Ganapathi, J. (2007). Putting the S back in corporate social responsibility: A multilevel theory of social change in organizations. ***Academy of Management Review***, *32*(3), 836–863.

Ahmadjian, C. L., & Robinson, P. (2001). Safety in numbers: Downsizing and the deinstitutionalization of permanent employment in Japan. ***Administrative Science Quarterly***, *46*(4), 622–654.

Allen, E. J., Dechow, P. M., Pope, D. G., & Wu, G. (2016). Reference-dependent preferences: Evidence from marathon runners. ***Management Science***.

Angrist, J. D., & Pischke, J.-S. (2008). ***Mostly harmless econometrics: An empiricist's companion***. Princeton University Press.

Avery, D. R. (2003). Reactions to diversity in recruitment advertising--are differences black and white? ***Journal of Applied Psychology***, *88*(4), 672.

Avery, D. R., & McKay, P. F. (2006). Target practice: An organizational impression management approach to attracting minority and female job applicants. ***Personnel Psychology***, *59*(1), 157–187.

Bainbridge, S. M., & Henderson, M. T. (2014). Boards-R-Us: reconceptualizing corporate boards. *Stan. L. Rev.*, *66*, 1051.

Bell, J. M., & Hartmann, D. (2007). Diversity in everyday discourse: The cultural ambiguities and consequences of "happy talk." ***American Sociological Review***, *72*(6), 895–914.

Bolino, M. C., Kacmar, K. M., Turnley, W. H., & Gilstrap, J. B. (2008). A multi-level review of impression management motives and behaviors. ***Journal of Management***, *34*(6), 1080–1109.

Brammer, S., & Millington, A. (2006). Firm size, organizational visibility and corporate philanthropy: An empirical analysis. ***Business Ethics: A European Review***, *15*(1), 6–18.

Brands, R. A., & Fernandez-Mateo, I. (2016). Leaning Out: How Negative Recruitment Experiences Shape Women's Decisions to Compete for Executive Roles. ***Administrative Science Quarterly***, 0001839216682728.

Buckley, C. (2016, January 15). Another Oscar Year, Another All-White Ballot. ***New York Times***. Retrieved from https://www.nytimes.com/2016/01/16/movies/oscar-ballot-is-all-white-for-another-year.html

Campbell, J. L. (2007). Why would corporations behave in socially responsible ways? An institutional theory of corporate social responsibility. ***Academy of Management Review***, *32*(3), 946–967.

Castilla, E. J., & Benard, S. (2010). The paradox of meritocracy in organizations. ***Administrative Science Quarterly***, *55*(4), 543–676.

Chiu, S.-C., & Sharfman, M. (2011). Legitimacy, visibility, and the antecedents of corporate social performance: An investigation of the instrumental perspective. ***Journal of Management***, *37*(6), 1558–1585.

Cialdini, R. B. (2003). Crafting normative messages to protect the environment. ***Current Directions in Psychological Science***, *12*(4), 105–109.

Cialdini, R. B. (2007). Descriptive social norms as underappreciated sources of social control. ***Psychometrika***, *72*(2), 263.

Cialdini, R. B., Kallgren, C. A., & Reno, R. R. (1991). A focus theory of normative conduct: A theoretical refinement and reevaluation of the role of norms in human behavior. ***Advances in Experimental Social Psychology***, *24*, 201–234.

Cialdini, R. B., Reno, R. R., & Kallgren, C. A. (1990). A focus theory of normative conduct: recycling the concept of norms to reduce littering in public places. ***Journal of Personality and Social Psychology***, *58*(6), 1015.

Cialdini, R. B., & Trost, M. R. (1998). Social influence: Social norms, conformity and compliance. Retrieved from http://psycnet.apa.org/psycinfo/1998-07091-021

Coffman, L. C., Featherstone, C. R., & Kessler, J. B. (2014). Can Social Information Affect What Job You Choose and Keep? ***American Economic Journal: Applied Economics***.

Desai, V. M. (2011). Mass media and massive failures: Determining organizational efforts to defend field legitimacy following crises. ***Academy of Management Journal***, *54*(2), 263–278.

Dezső, C. L., Ross, D. G., & Uribe, J. (2016). Is there an implicit quota on women in top management? A large□sample statistical analysis. ***Strategic Management Journal***, *37*(1), 98–115.

DiMaggio, P., & Powell, W. W. (1983). The iron cage revisited: Collective rationality and institutional isomorphism in organizational fields. ***American Sociological Review***, *48*(2), 147–160.

Dobbin, F., Schrage, D., & Kalev, A. (2015). Rage against the iron cage: The varied effects of bureaucratic personnel reforms on diversity. ***American Sociological Review***, *80*(5), 1014–1044.

Elsbach, K. D., & Sutton, R. I. (1992). Acquiring organizational legitimacy through illegitimate actions: A marriage of institutional and impression management theories. ***Academy of Management Journal***, *35*(4), 699–738.

Elsbach, K. D., Sutton, R. I., & Principe, K. E. (1998). Averting expected challenges through anticipatory impression management: A study of hospital billing. ***Organization Science***, *9*(1), 68–86.

Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, *7*(2), 117–140.

Fombrun, C. (1996). ***Reputation***. Wiley Online Library.

Fombrun, C., & Shanley, M. (1990). What's in a name? Reputation building and corporate strategy. ***Academy of Management Journal***, *33*(2), 233–258.

Forbes, D. P., & Milliken, F. J. (1999). Cognition and corporate governance: Understanding boards of directors as strategic decision-making groups. ***Academy of Management Review***, *24*(3), 489–505.

Gardberg, N. A., & Fombrun, C. J. (2006). Corporate citizenship: Creating intangible assets across institutional environments. ***Academy of Management Review***, *31*(2), 329–346.

Gino, F., & Pierce, L. (2010). Robin Hood under the hood: Wealth-based discrimination in illicit customer help. ***Organization Science***, *21*(6), 1176–1194.

Goldstein, N. J., Cialdini, R. B., & Griskevicius, V. (2008). A room with a viewpoint: Using social norms to motivate environmental conservation in hotels. ***Journal of Consumer Research***, *35*(3), 472–482.

Greve, H. R. (2008). A behavioral theory of firm growth: Sequential attention to size and performance goals. ***Academy of Management Journal***, *51*(3), 476–494.

Heath, C., Larrick, R. P., & Wu, G. (1999). Goals as reference points. ***Cognitive Psychology***, *38*(1), 79–109.

Highhouse, S., Brooks, M. E., & Gregarus, G. (2009). An organizational impression management perspective on the formation of corporate reputations. ***Journal of Management***, *35*(6), 1481–1493.

Kallgren, C. A., Reno, R. R., & Cialdini, R. B. (2000). A focus theory of normative conduct: When norms do and do not affect behavior. ***Personality and Social Psychology Bulletin***, *26*(8), 1002–1012.

Kanter, R. M. (1977). Some effects of proportions on group life: Skewed sex ratios and responses to token women. ***American Journal of Sociology***, 965–990.

King, B. G. (2008). A political mediation model of corporate response to social movement activism. ***Administrative Science Quarterly***, *53*(3), 395–421.

Konrad, A. M., Kramer, V., & Erkut, S. (2008). Critical Mass: The Impact of Three or More Women on Corporate Boards. ***Organizational Dynamics***, *37*(2), 145–164.

Leary, M. R., & Kowalski, R. M. (1990). Impression management: A literature review and two-component model. ***Psychological Bulletin***, *107*(1), 34.

Lee, J. C. (2017, January 13). Trump's Cabinet So Far Is More White and Male Than Any First Cabinet Since Reagan's. ***New York Times***. Retrieved from https://www.nytimes.com/interactive/2017/01/13/us/politics/trump-cabinet-women-minorities.html

Leslie, L. M., Manchester, C., & Dahm, P. (2017). Why and When Does the Gender Gap Reverse? Diversity Goals and the Pay Premium for High Potential Women. ***Academy of Management Journal***, *60*(2), 402–432.

Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. ***American Psychologist***, *57*(9), 705.

Matsa, D. A., & Miller, A. R. (2013). A female style in corporate leadership? Evidence from quotas. ***American Economic Journal: Applied Economics***, *5*(3), 136–169.

McDonnell, M. H., & King, B. (2013). Keeping up appearances reputational threat and impression management after social movement boycotts. ***Administrative Science Quarterly***, *58*(3), 387–419.

Merchant, N. (2013, October 7). Viewpoint: Twitter's All-Male Board Spells Failure. ***Time***. Retrieved from http://ideas.time.com/2013/10/07/viewpoint-twitters-all-male-board-spells-failure/

Miller, C. C. (2013, October 4). Curtain is rising on a tech premiere with (as usual) a mostly male cast. ***New York Times***. Retrieved from http://www.nytimes.com/2013/10/05/technology/as-tech-start-ups-surge-ahead-women-seem-to-be-left-behind.html

Miller, C. C. (2014, June 19). Women on the Board: Quotas Have Limited Success. ***New York Times***. Retrieved from https://www.nytimes.com/2014/06/20/upshot/women-on-the-board-quotas-have-limited-success.html

Moskowitz, T., & Wertheim, L. J. (2011). ***Scorecasting: The hidden influences behind how sports are played and games are won***. Crown Archetype. Retrieved from https://books.google.com/books?hl=en&lr=&id=0lSi_eCQvNwC&oi=fnd&pg=PA1&dq=moskowitz+wertheim+scorecasting&ots=Egl8R3BtiF&sig=E3TaC8aweoucW_5RyN6As6R5fCI

Nolan, J. M., Schultz, P. W., Cialdini, R. B., Goldstein, N. J., & Griskevicius, V. (2008). Normative social influence is underdetected. ***Personality and Social Psychology Bulletin***, *34*(7), 913–923.

Oliver, C. (1991). Strategic responses to institutional processes. ***Academy of Management Review***, *16*(1), 145–179.

Paluck, E. L., & Shepherd, H. (2012). The salience of social referents: A field experiment on collective norms and harassment behavior in a school social network. ***Journal of Personality and Social Psychology***, *103*(6), 899.

Pierce, L., Snow, D. C., & McAfee, A. (2015). Cleaning house: The impact of information technology monitoring on employee theft and productivity. ***Management Science***, *61*(10), 2299–2319.

Pollock, T. G., & Rindova, V. P. (2003). Media legitimation effects in the market for initial public offerings. ***Academy of Management Journal***, *46*(5), 631–642.

Pope, D., & Simonsohn, U. (2011). Round numbers as goals evidence from baseball, SAT takers, and the lab. ***Psychological Science***, *22*(1), 71–79.

Prentice, D. A., & Miller, D. T. (1993). Pluralistic ignorance and alcohol use on campus: some consequences of misperceiving the social norm. ***Journal of Personality and Social Psychology***, *64*(2), 243.

Rubinstein, R. Y., & Kroese, D. P. (2011). ***Simulation and the Monte Carlo method*** (Vol. 707). John Wiley & Sons. Retrieved from https://books.google.com/books?hl=en&lr=&id=yWcvT80gQK4C&oi=fnd&pg=PT10&dq=rubinstein+kroese&ots=xosXA8fboK&sig=Yk6woEUxf0PHvWY5QDk7vauSVJ4

Ryan, P. (2016, February 2). #OscarsSoWhite controversy: What you need to know. ***USA Today***. Retrieved from https://www.usatoday.com/story/life/movies/2016/02/02/oscars-academy-award-nominations-diversity/79645542/

Schultz, P. W., Nolan, J. M., Cialdini, R. B., Goldstein, N. J., & Griskevicius, V. (2007). The constructive, destructive, and reconstructive power of social norms. ***Psychological Science***, *18*(5), 429–434.

Shaffer, L. S. (1983). Toward Pepitone's vision of a normative social psychology: What is a social norm? ***Journal of Mind and Behavior***.

Shemla, M., Meyer, B., Greer, L., & Jehn, K. A. (2014). A review of perceived diversity in teams: Does how members perceive their team's composition affect team processes and outcomes? ***Journal of Organizational Behavior***.

Sherif, M. (1936). The psychology of social norms.

Simon, M., & Houghton, S. M. (2003). The relationship between overconfidence and the introduction of risky products: Evidence from a field study. ***Academy of Management Journal***, *46*(2), 139–149.

Smale, A., & Miller, C. C. (2015, March 6). Germany Sets Gender Quota in Boardrooms. ***New York Times***. Retrieved from https://www.nytimes.com/2015/03/07/world/europe/german-law-requires-more-women-on-corporate-boards.html

S&P Dow Jones Indices. (2015). ***S&P U.S. Indices Methodology****, October 2015*. Retrieved from https://us.spindices.com/indices/equity/sp-500

Staats, B. R., Dai, H., Hofmann, D., & Milkman, K. L. (2016). Motivating process compliance through individual electronic monitoring: An empirical examination of hand hygiene in healthcare. ***Management Science****, 63*(5), 1563–1585.

Staw, B. M. (1991). Dressing up like an organization: When psychological theories can explain organizational action. ***Journal of Management***, *17*(4), 805–819.

Sutton, R. I., & Galunic, D. C. (1996). Consequences of public scrutiny for leaders and their organizational image and its management. ***Research in Organizational Behavior***, *18*, 201–250.

Swim, J. K., & Hyers, L. L. (1999). Excuse me—What did you just say?!: Women's public and private responses to sexist remarks. ***Journal of Experimental Social Psychology***, *35*(1), 68–88.

Tankard, M. E., & Paluck, E. L. (2017). The effect of a Supreme Court decision regarding gay marriage on social norms and personal attitudes. ***Psychological Science***, *28*(9), 1334–1344.

Torchia, M., Calabrò, A., & Huse, M. (2011). Women directors on corporate boards: From tokenism to critical mass. ***Journal of Business Ethics***, *102*(2), 299–317.

Totenberg, N. (2016, September 23). Donald Trump Unveils New, More Diverse Supreme Court Short List. *NPR*. Retrieved from http://www.npr.org/2016/09/23/495216645/donald-trump-unveils-new-more-diverse-supreme-court-short-list

Unzueta, M. M., & Binning, K. R. (2010). Which racial groups are associated with diversity? ***Cultural Diversity and Ethnic Minority Psychology***, *16*(3), 443–446.

Unzueta, M. M., & Binning, K. R. (2012). Diversity Is in the Eye of the Beholder How Concern for the In-Group Affects Perceptions of Racial Diversity. ***Personality and Social Psychology Bulletin***, *38*(1), 26–38.

Unzueta, M. M., Knowles, E. D., & Ho, G. C. (2012). Diversity Is What You Want It to Be How Social-Dominance Motives Affect Construals of Diversity. ***Psychological Science***, *23*(3), 303–309.

Zavyalova, A., Pfarrer, M. D., Reger, R. K., & Shapiro, D. L. (2012). Managing the message: The effects of firm actions and industry spillovers on media coverage following wrongdoing. ***Academy of Management Journal***, *55*(5), 1079–1101.

Link to *Online Supplement*: https://osf.io/562yg/?view_only=1d8c31b6e5a94b0aa40ee90ac95f3f5a

FIGURE 1 Comparison of Actual Distribution of Women on (A) S&P 1500 Boards and (B) S&P 500 Boards with Simulated Expected Distribution of Women

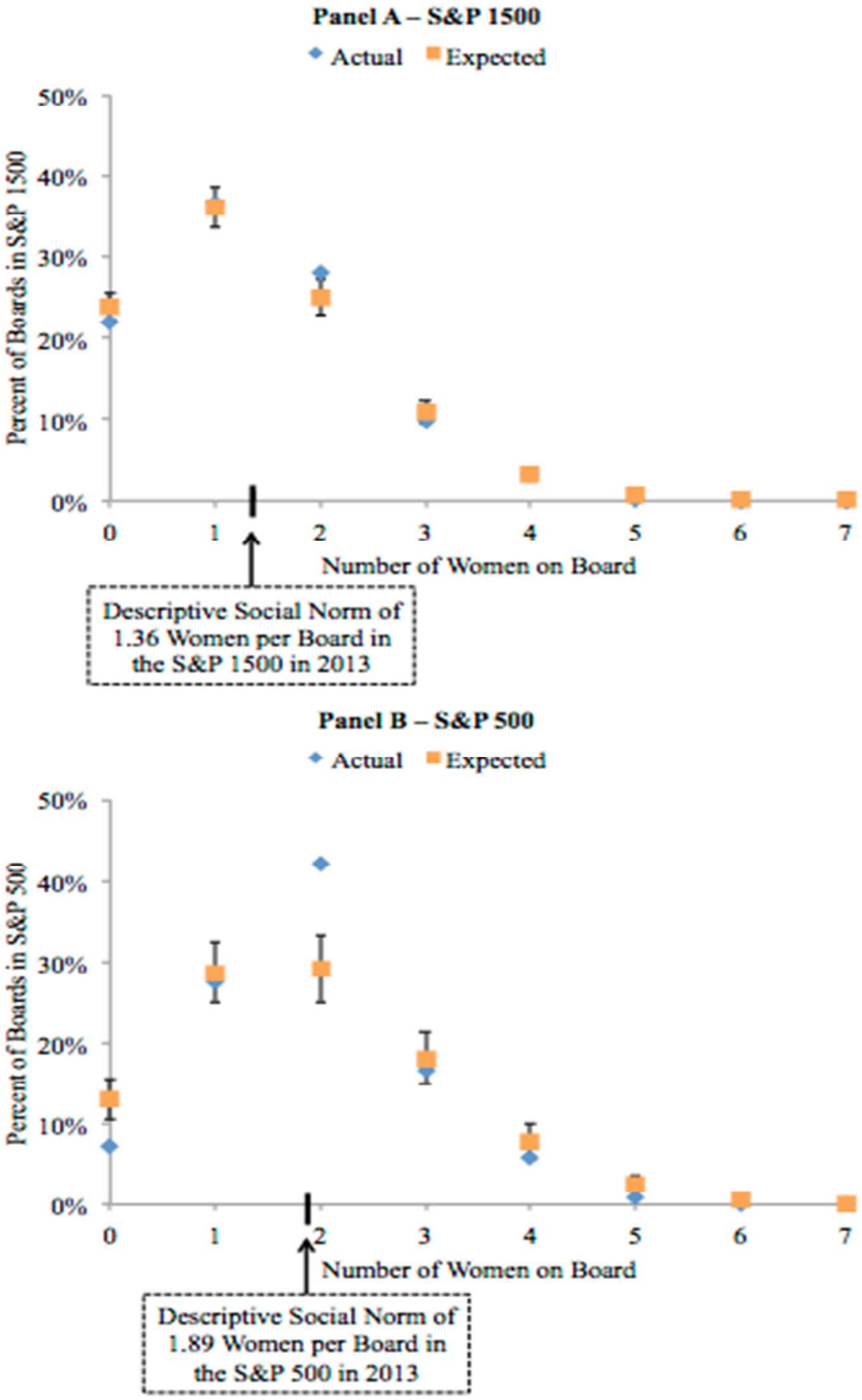
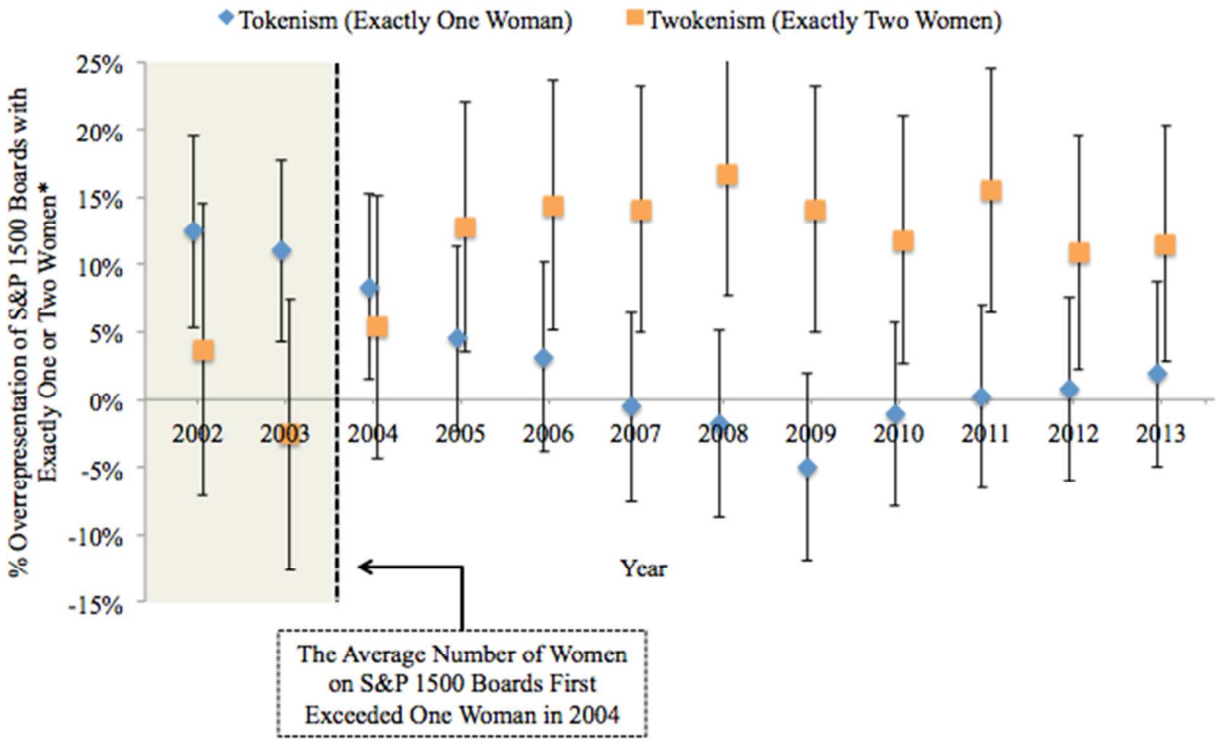
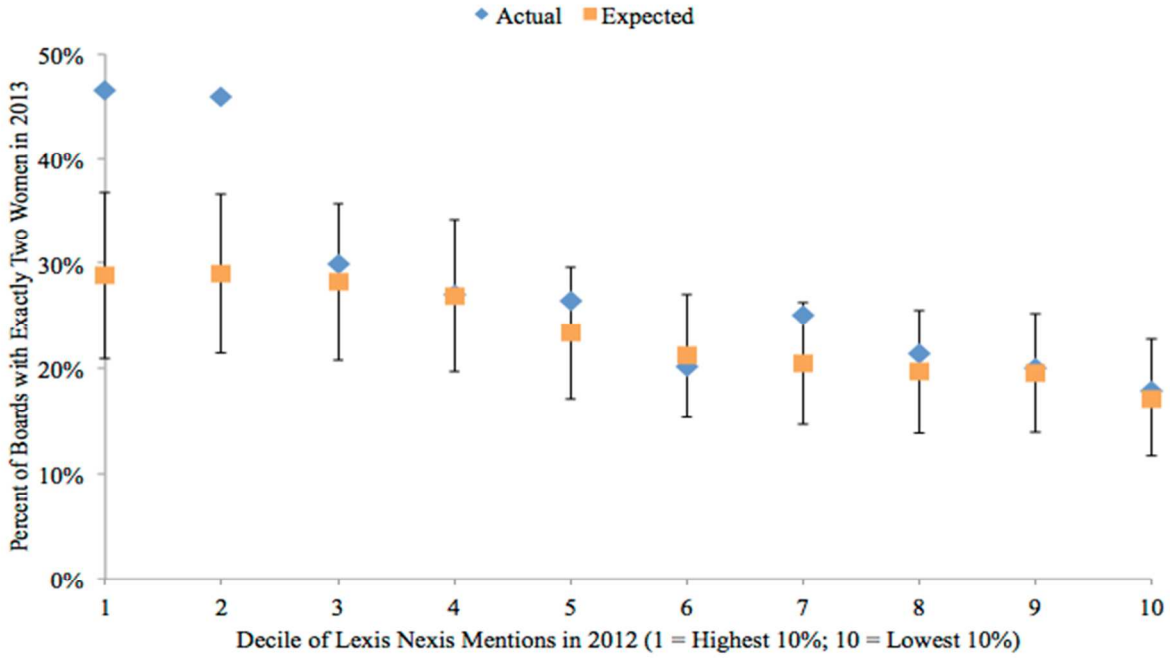


FIGURE 2 How Tokenism and Twokenism Shifted as Social Norms Changed from 2002 to 2013



* % Overrepresentation of S&P 1500 Boards with Exactly One Woman is calculated for each year using the following formula: [actual number of S&P 1500 boards with exactly one woman – expected number of S&P 1500 boards with exactly one woman] / [expected number of S&P 1500 boards with exactly one woman] where the expected number is calculated using our standard Monte Carlo simulation method. % Overrepresentation of S&P 1500 Boards with Exactly Two Women is calculated for each year using the following formula: [actual number of S&P 1500 boards with exactly two women – expected number of S&P 1500 boards with exactly two women] / [expected number of S&P 1500 boards with exactly two women] where the expected number is calculated using our standard Monte Carlo simulation method.

FIGURE 3 Firm Visibility Moderates the Extent of Twokenism



Note: This figure depicts the comparison between the actual percent of boards with exactly two women and the simulated expected percent of boards with exactly two women in 2013 as moderated by a firm's visibility. Firm visibility is operationalized using media mentions measured from Lexis Nexis, and firms are grouped into deciles based on the number of media mentions they received in 2012.

FIGURE 4 Participants in Study 1C Less Likely to Increase Gender Diversity of Boards Once Boards Include Two Women (and Thus Exceed the Social Norm)

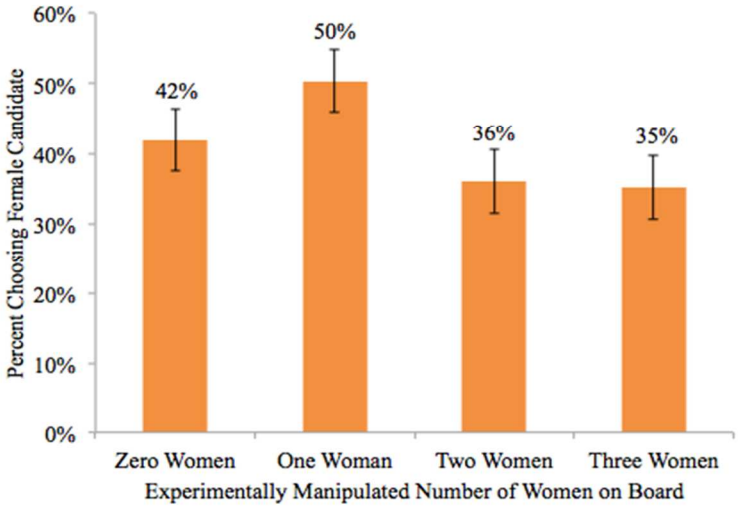


FIGURE 5A Participants' Preferences for Women Are Influenced by Social Norms and Scrutiny in Study 2A

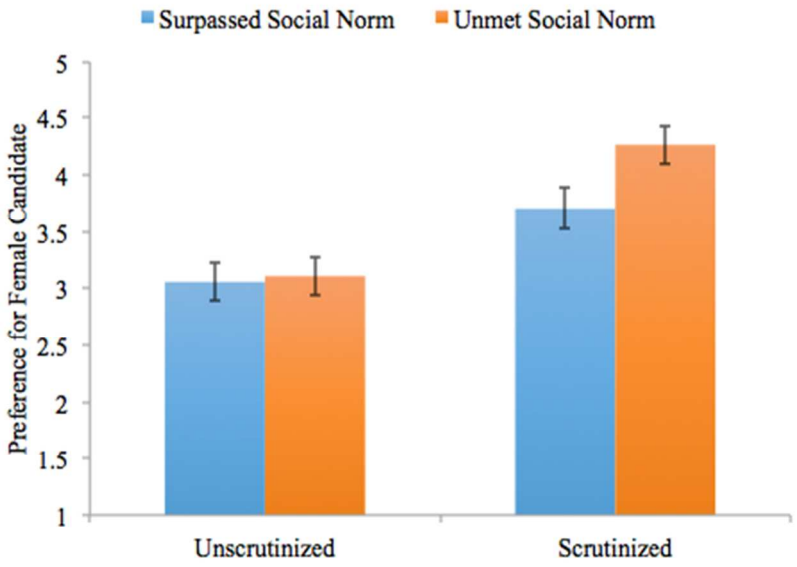


FIGURE 5B Interaction Between Social Norms and Scrutiny in Study 2B

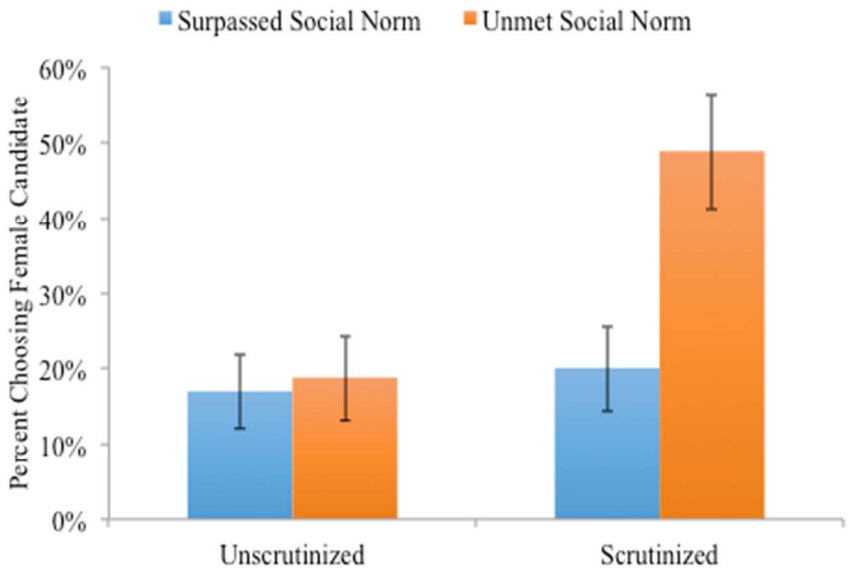


FIGURE 6 Interaction Between Social Norms and Visibility in Study 3

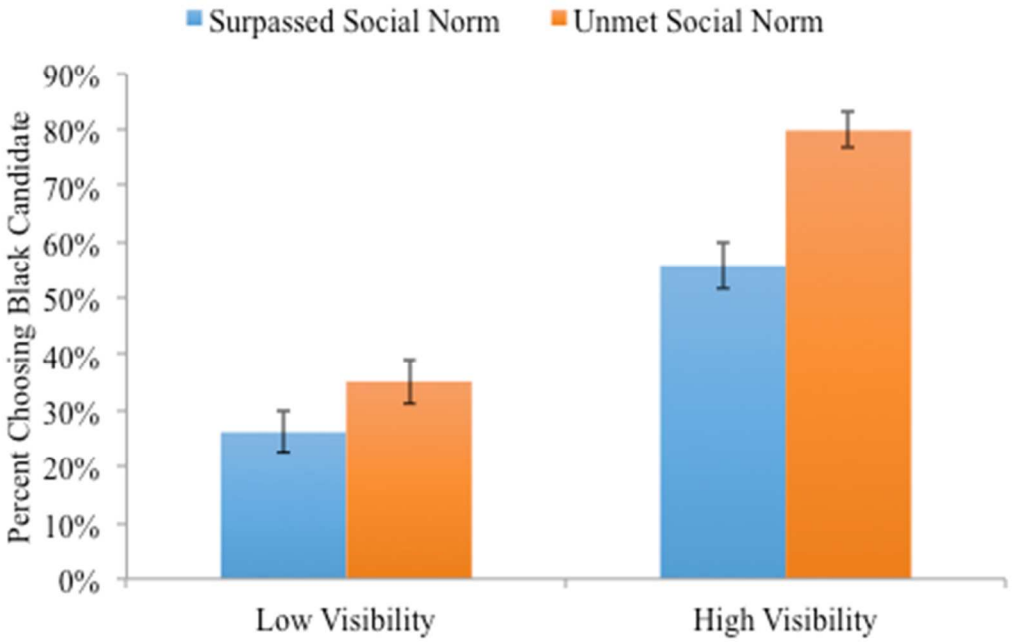


TABLE 1 Summary Statistics Describing S&P 1500 Dataset

	Proportion of all Directors
Male	86%
Female	14%
Caucasian	91%
Asian	3.0%
Black	3.7%
Hispanic	1.7%
Other Ethnicity	0.81%
1 Board Seat	84%
2 Board Seats	13%
3 Board Seats	2.8%
4 Board Seats	0.37%
5 Board Seats	0.07%

TABLE 2 Correlation Matrix for S&P 1500 Board Data in 2013 (N = 1,441)

	1	2	3	4	5	6
1. Size of Board	1.00					
2. Number of Female Directors	0.51***	1.00				
3. Number of Racial Minority Directors	0.36***	0.30***	1.00			
4. Logarithm of Market Capitalization	0.44***	0.36***	0.31***	1.00		
5. Logarithm of Media Mentions	0.43***	0.38***	0.32***	0.59***	1.00	
6. Member of S&P 500	0.43***	0.36***	0.29***	0.71***	0.59***	1.00

*, **, and *** denote significance at the 5%, 1%, and 0.1% levels, respectively

TABLE 3 Comparison of Actual and Expected Number of Female Directors Across S&P 1500 Boards of Different Sizes

Size of Board	*n*	Excess Percentage of Boards Observed with 0 Female Directors	Excess Percentage of Boards Observed with 1 Female Director	Excess Percentage of Boards Observed with 2 Female Directors	Excess Percentage of Boards Observed with 3 Female Directors
6 or fewer	124	2.74%	-5.79%	-16.96%	37.36%
		(2.89%)	(11.23%)	(28.15%)	(109.43%)
7	199	-2.60%	4.38%	10.06%	-66.06%
		(3.34%)	(8.98%)	(16.79%)	(54.18%)
8	241	-15.98%**	23.15%**	-8.42%	-24.81%
		(5.57%)	(8.04%)	(11.18%)	(24.50%)
9	283	-26.32%**	14.16%*	20.60%*	-34.31%*
		(8.05%)	(7.22%)	(9.09%)	(14.04%)
10	235	-38.00%***	18.88%*	16.51%	-16.12%
		(10.85%)	(8.58%)	(10.14%)	(13.57%)
11	198	-50.56%***	6.67%	28.69%**	7.47%
		(14.71%)	(9.95%)	(10.95%)	(13.42%)
12	100	-76.36%**	-22.97%	64.40%***	-6.40%
		(28.08%)	(15.09%)	(15.38%)	(17.34%)
13 or more	134	-58.37%*	-20.70%	49.34%***	22.85%
		(24.01%)	(13.63%)	(13.10%)	(15.55%)

Note: This table reports the difference between the actual percent of boards with a given number of female directors and the simulated expected percent of boards with that number of female directors conditional on the size of the board. Standard deviations are reported in parentheses.

*, **, and *** denote significance at the 5%, 1%, and 0.1% levels, respectively

TABLE 4 Boards Less Likely to Add Additional Women Once They Include at Least Two Women

	Board Added Woman = 1 (ordinary least squares regression)			
	Model 1	**Model 2**	**Model 3**	**Model 4**
Sample:	**S&P 1500**	**S&P 1500**	**S&P 500**	**S&P 500**
Number of Women on Board	-0.0033	-0.039***	-0.0056	-0.035*
	(0.0079)	(0.0090)	(0.012)	(0.015)
Indicator for Two or More Women on Board	-0.039*	-0.034*	-0.092***	-0.090***
	(0.016)	(0.016)	(0.023)	(0.024)
Controls Present	No	Yes	No	Yes
Observations	9,989	9,936	4,131	4,117
R^2	0.0032	0.030	0.017	0.045

Note. This table shows a series of ordinary least squares (OLS) regressions predicting whether boards add women conditional on the number of women already on the board and whether the board had met the descriptive social norm for gender diversity (i.e. already had at least two women) in the S&P 1500 (Models 1 and 2) and the S&P 500 (a subset of the S&P 1500; Models 3 and 4). Robust standard errors are in parentheses. When controls are present, regressions include fixed effects for board size, fixed effects for industry, fixed effects for stock market index, and a continuous control for market capitalization.
*, **, and *** denote significance at the 5%, 1%, and 0.1% levels, respectively

TABLE 5 More Visible Companies Show Larger Discontinuities at the Descriptive Social Norm

	Board Added Woman = 1 (ordinary least squares regression)	
	Model 1	**Model 2**
Number of Women on Board	-0.017*	-0.043***
	(0.0083)	(0.0090)
Indicator for Two or More Women on Board	-0.27	-0.023
	(0.017)	(0.017)
Centered Logarithm of Media Mentions	0.026***	0.017**
	(0.0042)	(0.0049)
Number of Women on Board x Centered Logarithm of Media Mentions	-0.0018	-0.00018
	(0.0049)	(0.0053)
Indicator for Two or More Women on Board x Centered Logarithm of Media Mentions	-0.021*	-0.021*
	(0.010)	(0.010)
Controls Present	No	Yes
Observations	9,781	9,743
R^2	0.012	0.033

Note. This table shows two ordinary least squares (OLS) regressions predicting whether boards add women conditional on the number of women already on the board and whether the board had met the descriptive social norm for gender diversity (i.e. already had at least two women), interacted with the centered logarithm of the number of media mentions a company receives. Robust standard errors are in parentheses. When controls are present, regressions include fixed effects for board size, fixed effects for industry, fixed effects for stock market index, and a continuous control for market capitalization.
*, **, and *** denote significance at the 5%, 1%, and 0.1% levels, respectively

TABLE 6 Regression Predicting the Selection of the Female Candidate to Serve on a Corporate Board in Study 1C

	B
Number of Women on Original Board	0.0401
	(0.0448)
Original Board Has Two or More Women	-0.187†
	(0.100)
Observations	479
R^2	0.0134

Note. This table shows the results of an ordinary least squares regression (OLS) predicting whether participants added a woman to a board conditional on the number of women already on the board and whether the board had met the descriptive social norm for gender diversity (i.e. already included at least two women). Robust standard errors are in parentheses.
†, *, **, and *** denote significance at the 10%, 5%, 1%, and 0.1% levels, respectively

TABLE 7
Regression Predicting Preference for Female Candidates to Serve on Panels in Studies 2A and 2B

	Model 1, Study 2A	**Model 2, Study 2B**
DV:	***Rating of Female Candidate***	***Chose Female Candidate***
Scrutinized	0.647**	0.031
	(0.243)	(0.075)
Unmet Social Norm	0.0511	0.018
	(0.244)	(0.075)
Scrutinized x Unmet Social Norm	0.508	0.270*
	(0.344)	(0.122)
Observations	556	200
R^2	0.0563	0.0837

Note. These ordinary least squares (OLS) regressions present the preference for the female candidate to serve on a panel in Studies 2A and 2B. Scrutinized is an indicator for the Scrutinized condition. Unmet Social Norm is an indicator for the unmet social norm condition. Robust standard errors are in parentheses.
*, **, and *** denote significance at the 5%, 1%, and 0.1% levels, respectively

TABLE 8 Regression Predicting the Selection of a Black Candidate for a Team in Study 3

	B
High Visibility	0.296***
	(0.055)
Unmet Social Norm	0.089
	(0.053)
High Visibility x Unmet Social Norm	0.151*
	(0.043)
Observations	603
R^2	0.175

Note. This ordinary least squares (OLS) regression predicts whether participants chose the black candidate to serve on a team in Study 3. High Visibility is an indicator for the High Visibility condition. Unmet Social Norm is an indicator for the unmet social norm condition. Robust standard errors are in parentheses.

*, **, and *** denote significance at the 5%, 1%, and 0.1% levels, respectively

Edward H. Chang (changed@wharton.upenn.edu) is a PhD Candidate at the Wharton School of the University of Pennsylvania. His research interests include diversity, discrimination, and behavior change.

Katherine L. Milkman (kmilkman@wharton.upenn.edu) is the Evan C Thompson Endowed Term Chair for Excellence in Teaching and an Associate Professor at the Wharton School of the University of Pennsylvania. Her research relies on field data and field experiments to understand the forces that shape people's decisions. She received her PhD in computer science and business from Harvard University.

Dolly Chugh (dchugh@stern.nyu.edu) is an Associate Professor in the Management and Organizations group at the New York University Stern School of Business. She received her PhD in Social Psychology and Organization Behavior from Harvard University. Her research interests include bounded ethicality and unconscious bias.

Modupe Akinola (makinola@columbia.edu) is the Sanford C. Bernstein & Co. Associate Professor of Leadership and Ethics at Columbia Business School. She received her PhD in organizational behavior from Harvard University. Her research explores how stress affects workplace performance. She also examines the strategies organizations employ to increase diversity, as well as the biases that affect the recruitment and retention of women and minorities in organizations.