

DO WOMEN FACE A HIGHER ETHICAL BAR?
EXPLORING GENDER DISCRIMINATION IN THE PUNISHMENT
OF ETHICAL VIOLATIONS AT WORK

Jessica A. Kennedy

Owen Graduate School of Management

Vanderbilt University

jessica.kennedy@owen.vanderbilt.edu

Mary-Hunter McDonnell

The Wharton School

University of Pennsylvania

marymcd@wharton.upenn.edu

Nicole M. Stephens

Kellogg School of Management

Northwestern University

n-stephens@kellogg.northwestern.edu

Please direct correspondence to:

Jessica A. Kennedy

401 21st Ave. S

Nashville, TN 37203

jessica.kennedy@owen.vanderbilt.edu

Acknowledgements

We thank Batia Wiesenfeld, Laura Kray, Sreedhari Desai, Frank Flynn, the Management of Organizations group at UC Berkeley, the Organisational Behaviour group at INSEAD, and three very constructive anonymous reviewers for their generous help with developing this manuscript. We thank Ryan Corser, James Northcutt, and Cooper Killen for their assistance with data collection.

ABSTRACT

Following an ethical violation at work, the punishments people receive have important implications for their lives, both professionally and personally, so it is vital that punishments be fair and non-discriminatory. However, across two studies using different methods, we find that women are punished more severely than men for ethical violations at work. Study 1 observes the gender disparity in an archival data set involving attorneys who violated the American Bar Association's ethical rules. Study 2 replicates the gender disparity in a laboratory experiment which randomly varies the gender of the transgressor. Our studies also explore both *when* and *why* this gender disparity emerges. To explore *when* this disparity emerges, we examined two potential moderating factors: the gender composition of the decision-making body that allocates punishment (Study 1) and the importance of ethicality to the group (Study 2). In Study 1, the gender disparity emerged only when women were under-represented among the decision-makers allocating punishment. In Study 2, the gender disparity emerged only when groups more strongly prescribed ethical traits. To explore *why* the disparity emerges, we examined the possibility that women face an intensified prescription to be ethical, relative to men. We found evidence for the prescriptive gender stereotype, but it did not mediate the gender disparity in punishment. Our research contributes to knowledge of punitive processes in work organizations and describes a novel form of gender discrimination.

Keywords: Gender; Ethical violations; Punishment; Stereotypes; Professionalism; Ethics; Misconduct

A large body of work highlights the impact of gender discrimination on allocative decisions in the workplace. Women receive fewer job offers (Goldin and Rouse, 2000), earn lower pay and fewer promotions (Belliveau, 2005, 2012; Castilla, 2008; Dreher & Cox, 2000; Fernandez-Mateo, 2009; Joshi, 2014; Olson & Becker 1983; Wright, Baxter, & Birkelund, 1995), secure less support for their entrepreneurial ventures (Brooks, Huang, Kearney, & Murray, 2014), obtain worse outcomes from negotiation (Ayes & Siegelman, 1995; Bowles, Babcock, & Lai, 2007), even when their behaviors, qualifications, and performance are identical to men's. Collectively, this research documents gender discrimination in the allocation of employment-related *rewards*. Here, we investigate a distinct but complementary question: Does the gender of employees shape the allocation of *punishment* for ethical violations at work?

Gender discrimination in the punishment of ethical violations at work is an important phenomenon for a number of reasons. First, ethical violations in professional settings are not uncommon. Incidents of fraud, improper relationships, conflicts of interest, and negligence occur across professions, including business, law, medicine, public service, engineering, law enforcement, and academia. According to one nationally representative survey, 41% percent of U.S.-based workers have observed ethical violations on the job (National Business Ethics Survey, 2013). And following ethical violations, transgressors typically incur punishment, defined as the application of an aversive event or the removal of a pleasing event in response to undesirable behavior (Arvey & Ivancevich, 1980). Each year in the United States alone, approximately 4,000 doctors are disciplined by state medical boards, over 900 attorneys are disbarred, 1,100 police officers are arrested, and fallen businesspeople and politicians publicly resign (American Bar Association, 2016; Federation of State Medical Boards, 2016; Stinson, Liederbach, Lab, & Brewer, 2016). Off the public record, employees face punishment in the form of docked bonuses, unfavorable assignments, and demotions. Thus, workplace punishment is a prevalent and important phenomenon to understand, as recognized by prior research (e.g., Wiltermuth & Flynn, 2013).

Second, the equitableness of workplace punishment is both theoretically and practically meaningful. In theory, punitive processes should be blind to demographic characteristics and determined

solely by the culpability of the transgressor and the harm caused by the transgression (Stanford Encyclopedia of Philosophy, 2015). In practice, however, researchers have only begun to understand the many personal, psychological, and situational factors that affect punitive intentions (Tetlock, 2002; Wiltermuth & Flynn, 2013) and invite unintentional bias into punitive processes (Albonetti, 1998; Nadler and McDonnell, 2012; Steffensmeier, Ulmer & Kramer, 1998). Workplace punishments also have important practical outcomes, as they can harm the careers and psychological well-being of recipients (Arvey & Ivancevich, 1980; Butterfield, Trevino, Wade & Ball, 2005; Trevino, 1992). When punished unfairly, people's job performance, commitment, and organizational citizenship behavior decline, while counterproductive behaviors increase (Cohen-Charash & Spector, 2001). Accordingly, one essential task of those who mete out punishment in professional settings is to do so fairly, without bias (Tetlock, 2002; Tetlock et al., 2007).

While to our knowledge no prior research has directly examined how gender impacts punishment for ethical violations at work, there is reason to expect gender discrimination. Generally, women are at greater risk than men of receiving informal social punishments in the workplace. Women are more often disliked and harassed by others for behaving assertively (Berdahl, 2007; Bowles et al., 2007; Brescoll, 2011; Rudman & Phelan, 2008). Likewise, when women make small mistakes, they face a greater loss of social status (Brescoll, Dawson, & Uhlmann, 2010). Moreover, the leadership positions offered to women tend to be at companies that are performing poorly, setting them up to fail (Cook & Glass, 2014; Ryan & Haslam, 2007; Ryan, Haslam, Hersby, & Bongiorno, 2011). For these reasons, women's professional status is more precarious than men's. In the same vein, we theorize that harsher punishment for ethical violations could be an additional, undocumented obstacle that women must navigate to sustain professional success.

Our central hypothesis is that women are punished more severely than men for ethical violations at work. Taking a multi-method approach, we investigate the relationship between gender and workplace punishment across two studies. Study 1 examines the relationship through a field study of attorneys who

violated the American Bar Association's ethical rules. Next, Study 2 utilizes a randomized, controlled laboratory experiment to help establish the causal role of the target's gender in producing the gender disparity and to further explore when and why a gender disparity emerges. Our multi-method approach has significant advantages (Chatman & Flynn, 2005). Our field data establish that the phenomenon has meaningful consequences in actual professional settings, whereas our laboratory experiment provides clear evidence documenting the causal role of gender.

Our research makes a number of important theoretical contributions. First, we draw attention to and provide robust, multi-method evidence for a novel form of gender discrimination: disparate punishment for ethical violations at work. Both gender (Heilman, 2012) and morality (Goodwin, Piazza, & Rozin, 2014) scholars have theorized that ethical violations could provoke different social reactions depending on the transgressor's gender, but this idea has not yet received a direct empirical test. Second, we explore *when* this form of discrimination emerges, attending to two group-level factors as potential moderators: the gender composition of the decision-making body that allocates punishment and how strongly the group prescribes ethical traits. Third, we begin to explore *why* this form of discrimination emerges by examining whether women face an intensified prescription to be ethical relative to men. Finally, we extend knowledge of the factors that impact punitive responses toward people who commit ethical transgressions (cf. Alicke, 2000; Arvey & Ivancevich, 1980; Butterfield et al., 2005; Tetlock et al., 2007; Trevino, 1992; Wiltermuth & Flynn, 2013).

Gender and the Punishment of Ethical Violations at Work

Decision-makers typically have substantial leeway to determine punishment for ethical violations. Inside firms, the at-will employment doctrine provides a great deal of managerial discretion (Longenecker, Sims, & Gioia, 1987; Murg & Scharman, 1982). As a result, managers wield significant power over employees. For instance, managers can determine whether employees are promoted, fired, assigned to dead-end roles, or denied bonuses. Although decision-makers are expected to act equitably, this expectation alone does not ensure objectivity. In fact, gender discrimination is especially likely to

emerge when decisions are made under a guise of objectivity (Castilla, 2008; Castilla & Benard, 2010; Uhlmann & Cohen, 2005).

Many factors are known to influence the severity of punishment applied for transgressions, including the power (Mooijman, van Dijk, Ellemers, & van Dijk, 2015; Wiltermuth & Flynn, 2013), status (Bowles and Gelfand, 2010), and political ideology of the punisher (Tetlock et al., 2007), as well as the perceived character and motives of the accused (Carlsmith, Darley, & Robinson, 2002; Nadler & McDonnell, 2012). How an employee's gender affects punishment of ethical violations is an open question. On one hand, there are reasons to think that women are punished *less* severely than men. People generally view women positively (Eagly & Mladinic, 1989) and see them as well-intentioned (Fiske, Cuddy, Glick, & Xu, 2002), so women who transgress might receive the benefit of a doubt. Moreover, lower-status perpetrators are punished less harshly (Fragale, Rosen, Xu, & Merideth, 2009; Tetlock et al., 2007), and women are seen as holding lower status in society than men (Ridgeway, 2001). On the other hand, research within the workplace context has found that women are punished *more* severely than men for whistle-blowing (Rehg, Miceli, Near, & Van Scotter, 2008). Likewise, women are punished more severely for deviant behaviors such as by "going against the boss's orders" or "covering up workplace mistakes" (Bowles & Gelfand, 2010). While these findings do not provide a direct test of our theory because the actions being punished do not represent ethical violations per se, these studies do provide indirect, preliminary evidence for the possibility that women are punished more severely than men for committing ethical violations at work. Despite the competing possibilities, for reasons that we explain in greater detail below, we theorize that women are punished *more* (rather than less) harshly than men following an ethical violation. Accordingly, we propose:

Hypothesis 1. *Women are punished more severely than men for committing an ethical violation at work.*

Gender and the Prescription to be Ethical

To begin to understand *why* gender might affect punishment for ethical violations at work, we first examine the nature of sexist attitudes as described by existing theory. Then, we consider how

descriptive sexist attitudes might translate into a prescriptive gender stereotype related to harsher punishment for ethical violations. Prior research has connected both sexist attitudes and prescriptive gender stereotypes to volatile, punitive social reactions toward women (Burgess & Borgida, 1999; Glick & Fiske, 2001; Fiske et al., 2002; Rudman & Phelan, 2008).

Sexist attitudes have been studied in the context of ambivalent sexism theory (Glick & Fiske, 1996, 2001), which holds that people simultaneously harbor antipathy toward women while revering and cherishing them (Glick & Fiske, 1996). These polarized attitudes are rooted in descriptive gender stereotypes such as “women seek to gain power by getting control over men,” and “women, compared to men, tend to have a superior moral sensibility” (Glick & Fiske, 1996, p. 512). People who endorse the negative statements about women also tend to endorse the positive ones, and thus the attitudes exhibit a moderate to strong positive correlation (Glick & Fiske, 1996, 2001).

We reason that these descriptive sexist beliefs – both negative and positive – could lead people to prescribe ethical behavior more strongly for women than for men. For example, to the extent that people vilify women as seductive, manipulative, and evil, they may hold women to a higher ethical standard in order to protect society from women’s corrupting influences and to deter other women from succumbing to their “natural” unethical proclivities. At the same time, to the extent that people sanctify women as morally pure, they may hold women to a higher ethical standard and punish them more harshly for ethical transgressions due to their exaggerated outrage at a woman’s deviation from her expected moral propriety. The connection between descriptive and prescriptive gender stereotypes has a long history. In fact, John Stuart Mill (1869) noted the connection between descriptive beliefs about women and their subjugation, arguing, “As for moral differences...[women] are declared to be better than men...which must provoke a bitter smile since there is no other situation in life in which it is... considered quite natural and suitable that the better should obey the worse” (p. 142). We would expect that both antipathy

toward women and cherishing of women could lead to the same general outcome – an intensified prescription for women to be ethical and harsher punishment for women who commit ethical violations.¹

Hypothesis 2. *Women face an intensified prescription to be ethical, relative to men.*

Hypothesis 3. *The intensified prescription for women to be ethical explains the gender disparity in punishment for ethical violations at work.*

Gender Composition of the Decision-Making Body as a Moderating Factor

Next, we consider two possible moderating factors that may help to explain *when* the gender disparity in punishment for ethical violations is likely to emerge. We consider two factors at the group level because gender bias often arises from the combination of micro-level forces (e.g., stereotypes) and macro-level forces (e.g., group composition or values) (Desai, Chugh, & Brief, 2014; Duguid, Lloyd, & Tolbert, 2012; Ely, 1995; Kanter, 1977; Martell, Emrich, & Robinson-Cox, 2012).

First, one potentially important factor is the gender composition of the group that allocates punishment. Extrapolating from research on minority influence (Peterson & Nemeth, 1996) and racial diversity (Antonio et al., 2004; Sommers, 2006), we reason that greater gender diversity among decision-makers could lead the group to deliberate more thoroughly and effectively. As a result, gender-diverse groups might base punishment decisions on case-relevant factors to a greater degree, relying less on prescriptive gender stereotypes. If so, then greater gender diversity among decision-makers could mitigate a gender disparity in punishment for ethical violations at work.

Greater gender diversity among decision-makers could improve the quality of deliberation by enhancing decision-makers' sense of accountability for making fair decisions. Indirect evidence for this idea comes from Sommers' (2006) study of racial diversity in mock juries. In that study, racially homogeneous or diverse groups were charged with evaluating an African American defendant from a Court TV trial. The presence of racial diversity on a jury improved the quality of the decisions in two

¹ These theoretical possibilities are speculative and our data do not arbitrate between them. See the General Discussion (p. 21) for additional consideration of how hostile and benevolent sexism might relate to the prescriptive gender stereotype.

ways. It reduced the likelihood that Caucasian participants would rely on stereotypes and perceive the defendant as guilty prior to any deliberation. Diversity also improved the thoroughness and accuracy of jury deliberations. Specifically, the discussions were longer, White participants cited a greater number of case-relevant facts, and the group discussion included fewer uncorrected factual errors. Collectively, the results suggest that Caucasian jurors might have felt more accountable for their judgments in the presence of an African American juror (cf. Lerner & Tetlock, 1999). In a similar vein, Antonio et al. (2004) found that racial and opinion diversity in groups enhanced peoples' ability to recognize alternative perspectives on an issue, increasing the complexity of their opinions. Together, the evidence suggests that diverse groups deliberate more effectively than homogenous ones.

Like other forms of diversity, gender diversity could improve decision-makers' deliberation process. If gender diversity, like racial and opinion diversity, promotes closer attention to the case facts and greater consideration and integration of alternative perspectives (Antonio et al. 2004; Sommers, 2006), then the punishments meted out by diverse groups might be determined by case-relevant factors (e.g., the nature of the ethical violation at issue and the harm it caused, the transgressor's intentions and history of behavior, and the broader context of the event), rather than by the moral outrage evoked by the violation of a prescriptive gender stereotype (Okimoto & Brescoll, 2010). Then, gender bias in punitive decisions could be reduced or eliminated altogether. On the basis of this logic, we propose:

Hypothesis 4. *The gender composition of the group that determines punishments for ethical violations at work moderates the extent to which women are punished more severely than men, such that the gender disparity is smaller when the group is more gender-balanced.*

How Strongly v. Weakly Groups Prescribe Ethics as a Moderating Factor

We next consider a second factor that could explain *when* a gender disparity in punishment for ethical violations emerges: namely, how strongly a group prescribes ethics. It is well-known that people and groups vary in how strongly they prescribe ethics, although existing research has largely tried to explain why this variation exists rather than examining its consequences (e.g., Cohen, Gunia, Kim-Jun, & Murnighan, 2009; Jackall, 2010; Kennedy & Anderson, 2017; Mayer et al., 2013). Building on prior

work, we expect that the strength of the prescription to be ethical will be an important predictor of when the gender disparity in punishment for ethical violations emerges.

We expect to find a gender disparity in punishment for ethical violations only when groups strongly prescribe ethics. Our reasoning is as follows: If groups strongly prescribe ethics (that is, if ethics are important to the group), then women should face the intensified prescription to be ethical. The concern about women maintaining ethical standards should only emerge when a group strongly endorses the idea that ethics are important. On the other hand, if groups weakly prescribe ethical traits (that is, do not think ethics are important), then it should not matter whether anyone exhibits ethical behavior, regardless of the person's gender. Instead, when groups weakly prescribe ethical traits, we expect that the gender disparity in punishment for ethical violations will disappear altogether. Therefore, we propose:

Hypothesis 5. *How strongly groups prescribe ethics moderates the extent to which women are punished more severely than men for ethical violations at work, such that the gender disparity emerges only when groups strongly prescribe ethical traits.*

STUDY 1: DISCRIMINATORY PUNISHMENT IN ATTORNEY DISCIPLINARY HEARINGS

Study 1 tested for a gender disparity in punishment for ethical violations at work (Hypothesis 1) and for a moderating role of the decision-making body's gender composition (Hypothesis 4). To examine these hypotheses in a field setting, we created an original database of archival decisions rendered in attorney disciplinary hearings, examining the severity of punishments applied to attorneys who violated the American Bar Association's (ABA) rules of ethical conduct.

This setting is ideal for testing our theory for a number of reasons. Punishment decisions in this context have important implications for attorneys' careers. Moreover, the decision-makers who allocate punishment are highly trained and they receive explicit guidance regarding how the rules should be applied and how to consider aggravating and mitigating factors when setting punishments. The context thus provides a conservative test of our hypotheses, as bias should be least likely to emerge when decision-makers are well-trained and the process is highly structured. In contrast to many settings in which punishments are conducted in private and documented unsystematically, the punishments allotted

by attorney disciplinary hearings are publicly available. Further, what constitutes an ethical violation is unusually clear in this setting. The ABA's ethical rules are carefully defined and familiar to attorneys. In fact, admission to a state's bar requires attorneys to pass an exam that tests their knowledge of the ABA's Model Rules of Professional Conduct (MRPC). Reports of disciplinary hearings explicitly document each rule that the attorney is accused of violating, allowing for a controlled comparison of punishment across different types of ethical violations. Finally, compared to other settings, the severity of punishment applied in these cases is indexed in a clear and straightforward way, such that it is very clear which forms of punishment are more severe (e.g., disbarment) than others (e.g., probation).

Sample

We created a database of attorney disciplinary cases heard by all state-level appellate courts in the U.S. in the year 2008, as published in Westlaw's ALLSTATES database of legal cases. Westlaw is a comprehensive legal resource that includes a searchable database of decided cases. For our study, we selected all cases where the current version of the MRPC applied, which made it possible to uniformly control for the precise ethical rules that were violated in each case. The MRPC, written and promulgated by the ABA in 1983, codifies the duties of the legal profession. Although states are under no express obligation to adopt the MRPC, thirty-eight states had adopted the most recent version in full by 2008. Our final sample includes 482 cases heard across 33 states. The states with the most cases in our sample are New Jersey (111 cases), Colorado (59 cases), Indiana, and Kansas (each with 35 cases).

The MRPC includes a set of uniform standards to govern the imposition of punitive sanctions for attorneys who violate the rules. These standards are intended to promote consistency in the application of punishment for ethical violations across jurisdictions. Through the MRPC, the ABA provides a list of possible sanctions for violating the rules of ethical conduct, as well as general standards for when each sanction is appropriate. Possible sanctions include, in order of increasing severity: admonition (or private reprimand),² public reprimand (declaration of the impropriety of the attorney's action, without

² Because private reprimands are not publicly disclosed and our data set was generated through public records, we could not include this type of sanction in our analyses.

interrupting the right to practice), probation (a period of increased scrutiny, without interrupting the right to practice), suspension (a temporary interruption in the lawyer's right to practice), and disbarment (extinguishing the transgressor's status as an attorney).

Measures

Punishment severity. Our dependent variable, *punishment severity*, is an ordinal variable. It reflects different categories of disciplinary outcomes using the levels specified by the ABA. Specifically, the variable is coded "0" if the attorney was publically reprimanded or censured, "1" if the attorney was put on probation, "2" if the attorney was suspended for a discrete amount of time, "3" if the attorney was indefinitely suspended, and "4" if the attorney was disbarred. In cases where more than one category of punishment was applied, this variable reflects the harsher punishment. For instance, if an attorney was suspended for six months to be followed by a year of probation, their punishment was coded as "2."

Attorney's gender. Gender was discerned from the name and pronoun ("he" or "she") used to refer to the attorney in the text of the disciplinary hearing summaries reported in Westlaw. The variable was coded either "0" (for men) or "1" (for women). Our sample included 83 cases against female attorneys, which represented 17% of the cases overall.

Percentage of women on the judges' panel. Each hearing was overseen by a panel of judges. To explore the impact of the gender composition of the judges' panel on punishment, we included the percentage of women sitting on the judges' panel as both a control variable and as a moderator. The list of judges involved in the hearing was reported in over half of the cases in our sample. We identified each judge's gender by matching their names to their pictures and biographies on court websites.

Control variables. We control for a number of other factors which could create a spurious relationship between gender and punishment severity. First, to control for the *type of ethical violations* committed, we include fixed effects for each individual rule – as codified in the MPRC – that was found to be violated. Our sample of cases involves 39 discrete rules that were violated. Each rule was coded "1" if violated in the case and "0" if not. The ethical rules most often violated include Rule 8.4, Ethical

violations ($n = 313$); Rule 1.3, Diligence ($n = 186$); Rule 1.4, Communication ($n = 181$); Rule 1.15, Safekeeping Property ($n = 141$); and Rule 8.1, Bar Admission and Disciplinary Matters ($n = 114$). To avoid overestimation of our models, we control only for rules with at least 15 observations, thereby appearing in at least 3% of the full sample of cases. Second, we controlled for the *total number of rules violated* in a given case. This count variable ranged from a low of 1 to a high of 13 rule violations. Third, to account for the possibility that the size of the judges' panel impacted the punishment's severity, we controlled for the *total number of judges on the hearing panel*.

Methods

Because our dependent variable, punishment severity, is an ordinal variable, we employed ordered logistic regression using the *ologit* command in Stata 14. This model essentially allows for the application of the standard logit model – which is appropriate for dichotomous outcome variables – to dependent variables which consist of more than two categories that have a natural underlying order to them. As a robustness check, we replicate our results with a probit model predicting the likelihood of the most severe punishment: disbarment. There, the dependent variable was coded “1” for cases in which an attorney was disbarred and “0” for cases ending in a less severe form of punishment.

To control for potential heteroskedasticity, we employ robust standard errors across all models, clustering by the state in which the case took place (Nordgren & McDonnell, 2011). Clustering standard errors in this manner takes into account the likelihood that state courts will afford greater precedential value to their own past cases than to cases held in other states. If so, disciplinary outcomes will be correlated more closely within-state than between-states. As an additional control for state-specific dynamics that could systematically affect punishments, we include a fixed effect for state. To avoid overestimation of our models, we control only for states with at least 15 observations, thereby appearing in at least 3% of the full sample of cases.

Results and Discussion

Table 1 provides descriptive statistics and correlations among variables.

In this study, we asked: Are women punished more severely than men for ethical violations in their profession (Hypothesis 1), and does the degree of any gender disparity in punishment for ethical violations depend on the gender composition of the group allocating punishment (Hypothesis 4)? Results are shown in Table 2. To provide the first test of Hypothesis 1, Model 1 includes all 482 cases. Then, Models 2-5 add controls for the size of the judges' panel and the percentage of women on the judges' panel. Because only a subset of cases disclosed the judges empaneled for the hearing, the number of observations in Models 2-5 ($n = 270$) is lower than in Model 1. To test Hypothesis 4, Models 3 and 5 include the interaction between the attorney's gender and the percentage of female judges on the panel.

Punishment (Hypothesis 1). We find robust support across all models for our prediction (Hypothesis 1) that women are punished more severely than men for committing ethical violations at work. In Models 1-3, a significant positive association between gender and punishment severity emerges. This finding suggests that, controlling for the type of ethical violations committed, female attorneys are generally assigned harsher punishments for their transgressions, relative to males. This finding is replicated in the probit models depicted in Models 4-5, which suggest that women are more likely than men to be disbarred, controlling for the type of ethical violations committed. Post-estimation margins analysis of Model 4 with all control variables held to their means indicates that women have a 106% higher likelihood of being disbarred than men (at .35 and .17, respectively).

Gender composition of the decision-making body as a moderator (Hypothesis 4). Hypothesis 4 also received support, as the interaction between the attorney's gender and the percentage of women on the judges' panel is negative and statistically significant in both Models 4 and 5. This finding indicates that discriminatory punishment is mitigated when women are more equally represented on the decision-making body that allocates punishment. To assist with the interpretation of this finding, Figure 1 plots the predicted values derived from post-estimation margins analysis of Model 3. As the figure shows, when the percentage of women on the judges' panel was one standard deviation below the mean, a woman's likelihood of being disbarred is more than twice that of a man. However, the likelihood of being disbarred

is nearly identical between men and women when the percentage of women on the judges' panel is one standard deviation above the mean. Therefore, more equal representation of women among decision-makers allocating punishment reduced gender disparities in punishment for ethical violations in this context.

STUDY 2: DISCRIMINATORY PUNISHMENT IN A LABORATORY EXPERIMENT

Study 2 was designed to extend knowledge in four respects. First and foremost, Study 2 was designed to replicate the findings from Study 1 using an experimental design. By manipulating a professional's gender in a hypothetical scenario while holding other factors constant, the experimental approach helps to establish that it is the gender of the transgressor rather than other factors that leads to harsher punishment (Hypothesis 1). Second, Study 2 also tests for the existence of an intensified prescription for women to be ethical (Hypothesis 2) and examines whether this variable can help to explain the gender disparity in punishment for an ethical violation (Hypothesis 3). Third, Study 2 explores the impact of how strongly groups prescribe ethics in an effort to further understand when the gender disparity in punishment might emerge (Hypothesis 4). Finally, this study generalizes our findings to a new professional context: business rather than law.

Methods

Sample. Participants ($N = 232$) were either undergraduates enrolled in a business course at a southeastern university ($n = 150$) or MBA students at a northeastern university ($n = 82$). The undergraduates completed the lab study for course credit, whereas the MBA students completed it as part of a class exercise. We included both groups in the study in order to maximize our sample size. Participants believed the study to examine "decision-making in groups." Upon arrival at the classroom or laboratory, they were assigned to a group ($N = 76$) of three ($n = 72$) or four ($n = 4$) people. Students completed the task in groups in order to emulate the data from our field setting and to allow us to explore the prescriptive gender stereotype at the group level. Thirty-one participants (13%³) expressed suspicion

³ The measure of suspicion was highly sensitive and could over-estimate its occurrence. At the end of the final survey, participants were asked, "What do you think this study is about?" Anyone who said something about

that the study examined “gender” and were excluded from analyses, leaving a final sample of 201 participants (44% women) ranging in age from 18 to 34 years ($M = 21.81$, $SD = 3.98$) and comprising 57 groups.

Design and procedure. The study utilized a two-condition, between-participants design (target’s gender: male, female). The gender composition of the group varied naturally, resulting in 35 mixed-gender groups, 16 all-male groups, and 6 all-female groups. Sitting with their group, participants read the following scenario:

Jake / Jill Moranty is employed as the lead manager of Bethany Central Hospital in Kansas City, Missouri. Moranty has worked at the hospital for twenty years. One of Moranty’s employment duties is to fill out the monthly Medicare reimbursable claims report for all surgeries performed in the emergency ward. Mr. / Ms. Moranty fills out the paperwork and then, once the Medicare office receives the report, it pays the hospital the amount owed for the reported services. This month, Moranty filed an erroneous claims report. In it, he / she intentionally reported several of the procedures twice, ultimately asking for \$10,000 more than what the hospital was owed. The Medicare office caught the discrepancy in the figures and has accused the hospital of Medicare fraud. The event has triggered an enormous amount of bad press for the hospital, and the hospital ultimately had to pay \$100,000 to the federal government to settle the charges.

This scenario has three key strengths. First, it describes an ethical violation that is non-relational, meaning that it occurs outside the context of a close, personal relationship (Dawson, 1992; Franke, Crown, & Spake, 1997). Second, the protagonist’s behavior is pro-organizational rather than self-interested (Vadera & Pratt, 2013). Third, it addresses dishonesty. In each respect, the ethical violation differs from a warmth violation. Ethical violations that undermine a close relationship (Dawson, 1992) or that benefit oneself (rather than the organization) (Amanatullah & Morris, 2010) could violate gender prescriptions of warmth. In contrast, dishonest acts have implications for morality but not for warmth (Goodwin et al., 2014). By examining a non-relational, pro-organizational act of dishonesty, we avoid confounding ethicality and warmth and provide a conservative test of our hypotheses.

“gender” was coded as suspicious. However, it is possible that participants erroneously believed the study was about the gender of decision-makers rather than the gender of the target. Most importantly, our results are virtually identical when all data are included in analyses. The woman ($M = 0.19$, $SD = 0.87$) is punished more harshly than the man ($M = -0.23$, $SD = 0.83$), $t(74) = 2.15$, $p = .035$, $d = 0.49$.

After reading this scenario, participants privately answered a survey that included measures of prescriptive gender stereotypes, detailed below. Then, the group discussed the scenario and collectively decided the appropriate level of punishment for the manager. Finally, they individually completed a survey with demographic questions and a question checking for suspicion.

Prescriptive gender stereotypes. To measure prescriptive gender stereotypes⁴, we closely followed the method of Prentice and Carranza (2002), asking how desirable or undesirable it would be for Jill / Jake to possess a number of characteristics using a scale from 1 (*very undesirable*) to 9 (*very desirable*). To measure a prescription to be ethical, we included eight (un)ethical traits drawn from existing research (Goodwin, 2015): cowardly, unfair, unprincipled, irresponsible, unjust, dishonest, untrustworthy, disloyal. Because people have been found to rate moral traits as very desirable (Aquino & Reed, 2002; Goodwin et al., 2014), the negatively-worded traits were used to help avoid ceiling effects. We reverse-scored the scale such that higher numbers indicated a stronger prescription to be ethical. The items were averaged to form a reliable scale ($\alpha = .91$).

Punishment intentions. Participants completed two items to indicate how severely they intended to punish the target. First, they reported how severely Jill / Jake should be punished for their ethical violation, using a scale from 1 (*not at all severely*) to 7 (*very severely*). Second, they read that Federal guidelines stipulate a fine for this type of crime ranging from 1x-10x the amount over-claimed and then indicated the appropriate fine for Jill / Jake, using a scale from 1x to 10x. The standardized items were highly correlated, $r(57) = .56, p < .001$, so we averaged them to form a scale ($\alpha = .72$).

Approach to analysis. Because one key goal in Study 2 was to explore the direction of the effect in Study 1, we continued to study punishment at the group level. Accordingly, we aggregated the measure

⁴ On an exploratory basis, we also measured prescriptions of competence ($\alpha = .90$; items drawn from Fiske et al., 2002), warmth ($\alpha = .89$; items drawn from Goodwin, 2015) and high-warmth ethical traits ($\alpha = .87$; items drawn from Goodwin, 2015). Competence was more intensely prescribed for the woman ($M = 8.47, SD = 0.44$) than for the man ($M = 7.98, SD = 0.63$) following the ethical violation, $t(55) = 3.27, p = .002, d = 0.89$. No significant differences by the target's gender emerged for the traits related to warmth ($ts < 1.66, ps > .10$), although the effect sizes suggest differences might emerge with a larger sample ($ds = 0.45$ and 0.37 for high-warmth ethical and non-ethical traits, respectively). Because these constructs are not the central focus of our investigation, we do not discuss them further.

of how strongly groups prescribed ethical traits to the group-level by forming an average on the construct for each of the 57 groups. Aggregating these ratings to the group level was justified on both conceptual and statistical grounds. Conceptually, this construct is a shared unit property, meaning that it originates in the characteristics of individual unit members but is likely to coalesce into a shared property through social interaction and other social processes (Kozlowski & Klein, 2000). Statistically supporting the aggregations, inter-rater agreement within the 57 groups was typically high (mean $r_{wg} = 0.97$).⁵

Results and Discussion

Because our data were collected at two sites, we first explored whether any significant differences emerged between the samples. No statistically significant differences in means emerged for the key variables ($ts < 0.40$, $ps > .69$, $ds < 0.14$). We therefore combined the samples in subsequent analyses. The data were analyzed using independent samples t -tests unless otherwise specified.

Punishment intentions (Hypothesis 1). We first explored Hypothesis 1, asking: Did participants intend to punish the woman more severely than the man for committing the ethical violation at work? Supporting Hypothesis 1, and generalizing results to a new context, participants intended to punish the woman ($M = 0.14$, $SD = 0.90$) more severely than the man ($M = -0.32$, $SD = 0.79$), $t(55) = 2.07$, $p = .044$, $d = 0.54$.

Prescriptive gender stereotypes (Hypothesis 2). Next, to test Hypothesis 2, we explored whether the woman faced an intensified prescription to be ethical, relative to the man. Supporting Hypothesis 2, ethical traits were seen as more desirable for the woman ($M = 8.58$, $SD = 0.35$) than for the man ($M = 8.31$, $SD = 0.59$), $t(55) = 2.05$, $p = .045$, $d = 0.56$.

Mediation analysis (Hypothesis 3). Then, to test Hypothesis 3, we examined whether the intensified prescription for women to be ethical could explain why participants intended to punish the woman more harshly than the man for the ethical violation. A bootstrapping analysis of mediation

⁵ Overall, the percentage of groups with r_{wg} equal or greater to 0.7, the standard cut-point for denoting high versus low inter-rater agreement (LeBreton & Senter, 2008) was 98%.

(Preacher & Hayes, 2008) with 10,000 samples yielded a 95% confidence interval of [-.02, .23] for the indirect effect, indicating no support for Hypothesis 3.⁶

Moderation by the strength of the ethical prescription (Hypothesis 5). Next, to test Hypothesis 5, we asked whether the effect of a target's gender on punishment depends on how strongly the group prescribes ethical traits. Again, we expected a gender disparity in punishment for ethical violations to emerge only when groups are concerned about maintaining ethical standards—that is, only when the prescription to be ethical is relatively strong. To test for such a relationship, we predicted punishment severity in a linear regression analysis with the target's gender, the strength of the ethical prescription, and the resulting interaction term. All explanatory variables were mean-centered prior to the analysis.

Results appear in Table 2. Within the full model, one main effect and an interaction emerged. First, groups that more strongly prescribed ethical traits punished the target more severely, $\beta = .40$, $t(53) = 2.41$, $p = .02$. Second, and most important for our purposes, a target gender X ethical prescription interaction emerged, $\beta = .35$, $t(53) = 2.18$, $p = .03$. To understand the source of this interaction, we plotted the results. Figure 3 depicts punishment as a function of the ethical prescription's strength (weak v. strong) and the target's gender. The ethical prescription categories (weak v. strong) were based on a median split, which facilitates interpretation (for a recent discussion of this method, see Iacobucci, Posavac, Kardes, Schneider, & Popovich, 2015). That is, groups received a 1 ($n = 29$) if their mean ethical prescription score was greater than the median on that measure and a 0 ($n = 28$) otherwise. Using the resulting binary variable for the ethical prescription's strength, we then examined the simple comparisons as a formal test of whether these observed differences were statistically significant. The simple comparisons revealed that, when the ethical prescription was strong, the target's gender influenced punishment severity at a statistically significant level, $F(1, 53) = 3.78$, $p = .02$, $\eta_p^2 = .18$. Specifically, the

⁶ In response to the unexpected null result, we explored whether the other intensified prescription found here – for competence – could help to explain the gender disparity in punishment for ethical violations. A separate bootstrapping analysis of mediation (Preacher & Hayes, 2008) with 10,000 samples yielded a 95% confidence interval of [.03, .38] for the indirect effect, providing evidence of significant mediation. However, given that this result is difficult to explain theoretically, we leave it for future research to explore.

female target ($M = 0.43$, $SE = 0.20$) was punished more severely than the male target ($M = -0.12$, $SE = 0.22$). When the ethical prescription was weak, the target's gender had no statistically significant effect on punishment severity, $F(1, 53) = 3.32$, $p = .07$, $\eta_p^2 = .06$. The findings supported Hypothesis 5.

Exploratory moderation by gender composition of the group (Hypothesis 4). Finally, to test Hypothesis 4, we explored whether the gender disparity in punishment for ethical violations was mitigated by more equal representation of women among the group allocating punishment, as found in Study 1. It is important to note that the current experiment was not designed to test this hypothesis, so this analysis was exploratory. That is, the primary goals of this study were to experimentally vary the target's gender and to provide a group context for making the punishment decision so that the prescription to be ethical could be explored at the group level. Since the key goal of the design was to identify the role of the target's gender and the prescription to be ethical in facilitating a gender disparity in punishment for ethical violations, the gender composition of the groups was not varied experimentally and variance along it was minimal. With these precautions noted, we proceed to test Hypothesis 4.

A linear regression analysis predicting punishment included mean-centered variables for the target's gender, the percentage of women in the group allocating punishment, and the target's gender X group composition interaction term. Only the variables for the target's gender, $\beta = .29$, $t(53) = 2.25$, $p = .028$, and the group composition, $\beta = .27$, $t(53) = 2.05$, $p = .046$, were statistically significant. The interaction term was non-significant, $\beta = .09$, $t(53) = 0.66$, $p = .52$. Therefore, Hypothesis 4 did not receive support in Study 2. Again, there are many possibilities as to why. The gender composition of the group varied naturally, resulting in many more mixed-sex ($n = 35$) than all-male ($n = 16$) groups, making the test under-powered statistically. An appropriately powered test would likely have required a sample size that is very difficult to obtain in the laboratory: between 200 and 600 groups of three people (i.e., 600 to 1,800 participants in order to obtain 50 to 150 groups per cell for a 4-condition design; Simmons, Nelson, & Simonsohn, 2018). Additionally, participants of both genders began the study in the same room, possibly exposing even all-male groups to gender diversity. Finally, the scenario method may not

have provided an ideal test of Hypothesis 4. While the approach provided excellent control by allowing us to change only the name and pronouns used to describe a target (enabling a strong and conservative test of Hypothesis 1), a central reason that more diverse groups could produce fairer outcomes is their tendency to more thoroughly and effectively process complex information (Antonio et al., 2004; Sommers, 2006). The limited information available through the scenario (relative to attorney disciplinary cases) may have constrained this mechanism. Thus, the null results must be interpreted with caution but do suggest a need for future work to explore the boundary effects of the gender composition moderator.

GENERAL DISCUSSION

It is vital for punishment to be distributed equitably in light of its implications for professionals' careers and psychological well-being. Yet we find evidence that a seemingly irrelevant factor—the gender of the transgressor—matters for how harshly workplace ethical violations are punished. These findings have important theoretical and practical implications.

Theoretical Contributions

Our research breaks new ground by documenting harsher punishment for ethical violations at work as a novel form of gender discrimination. It is well-known that professional women face social penalties for behaving in ways that benefit their careers, such as participating actively (Brescoll, 2011), displaying high levels of performance (Rudman & Fairchild, 2004), and negotiating their salaries (Bowles et al., 2007). Our research suggests that women also suffer disproportionate punishment for ethical violations at work. Thus, gender bias emerges not only in the allocation of employment-related rewards but also for employment-related punishment.

Our research also provides evidence for a new prescriptive gender stereotype: an intensified prescription for women to be ethical, relative to men. This stereotype might help to explain why women behave more ethically than men under some conditions (e.g., Cohen, Panter, Turan, Morse, & Kim, 2014; Eagly, Diekmann, Johannesen-Schmidt, & Koenig, 2004; Kennedy, Kray, & Ku, 2017). However, it failed to mediate the gender disparity in punishment for ethical violations.

Finally, we found preliminary evidence for two moderating factors. One potentially important factor is the gender composition of the group that allocates punishments.⁷ Greater representation can carry risks as well as benefits for women (McGinn & Milkman, 2013), but the results from Study 1 suggest that equal representation might lessen discrimination in the context of punishment for ethical violations at work. A second potentially important factor is how strongly the group prescribes ethical traits. It is non-intuitive that discrimination would emerge when groups more strongly prescribe ethics, but it is consistent with our notion that concern for maintaining ethical standards applies to women more than to men. We note that the results for both moderators are correlational and must therefore be interpreted with caution.

Practical Contributions

Our research makes practical contributions as well. First, it serves to warn professional women that unethical behavior is riskier for them than for their male colleagues; women could benefit from erring on the side of high ethical standards at work. Second, our findings could provide a beneficial practical suggestion for organizations that wish to avoid unfairly disadvantaging women—namely, to ensure that women are equally-represented on decision-making bodies that determine punishment for ethical violations.⁷ After all, responsibility for equitable punishment lies with those who mete it out.

Limitations and Future Directions

The primary strength of our research is its provision of robust, multi-method support for an important empirical phenomenon: a gender disparity in punishment for ethical violations at work. Future research is necessary to more fully unpack the psychological mechanisms underlying the phenomenon.

The new prescriptive gender stereotype we document here – the intensified prescription for women to be ethical, relative to men – also merits further exploration. It would be worth investigating the beliefs that underlie it. Earlier, we speculated that both hostile and benevolent sexism could undergird this prescription, but we were unable to empirically discern their relative contribution in our studies.⁸ Future

⁷ Given that we did not find that the gender composition of the decision-making body significantly moderated the extent of discriminatory punishment in our controlled experiment, this finding warrants further investigation.

⁸ We thank an anonymous reviewer for the suggestion to measure both perceived threat to the moral order (to approximate hostile sexism) and violated expectations (to approximate benevolent sexism). By doing so, we

research should empirically adjudicate between the potential theoretical explanations for it and explore its consequences.

Future research should also explore *how* a group's gender composition affects the extent of gender bias in its decisions. Although we focused on how gender composition might affect the thoroughness and accuracy of deliberations, it is unclear whether changes occur because women deliberate differently than do men, or whether men deliberate differently in the presence of women, or both.

Further work is required to identify boundary conditions for the gender disparity in punishment for ethical violations. One question is whether the disparity emerges outside of a professional domain. Preliminary evidence from a scenario study (see Study 3 in the Online Supplement) suggests that it might, but additional investigation is warranted. A second issue is whether the disparity emerges in contexts where women and men commit ethical violations at similar rates. Again, preliminary evidence suggests that it might (see Study 4 in the Online Supplement). A third issue is whether the type of ethical violation committed moderates the direction or degree of the gender disparity. It could be the case that women are punished less than men for ethical violations that advance others' interests, such as those that protect a friend from harm or obtain an unfair advantage for a child, as these are prototypically feminine goals (cf. Amanatullah & Morris, 2010; Amanatullah & Tinsley, 2013). Conversely, certain contexts could exacerbate discriminatory punishment. Women might be punished especially harshly for ethical violations that are intended to elevate their social status or assert dominance in a masculine domain, as punishment

examined whether descriptive gender stereotypes could explain the punishment disparity. Perceived threat to the moral order was measured using three items drawn from prior research (Fischer et al., 2007). Participants rated the extent to which Jill / Jake's crime was threatening to, dangerous for, and very problematic for our current society ($\alpha = .92$), using a scale from 0 (not at all) to 10 (definitely). Additionally, participants provided a measure of violated expectations using three items on the same scale: I am surprised that Jill / Jake committed this type of crime, I did not expect a person like Jill / Jake to commit this type of crime, and I thought that it was unlikely that Jill / Jake would commit this type of crime ($\alpha = .91$). We reasoned that greater hostile sexism (with its negative views of women) would lead people to perceive ethical violations by women to pose greater threat to the moral order, whereas greater benevolent sexism (with its positive views of women) would lead people to have violated expectations. Inter-rater agreement within the groups was high for both constructs (mean $r_{wg} \geq 0.85$), supporting aggregation of the constructs to the group level. However, no statistically significant effect of target's gender emerged for either perceived threat to the moral order, $F(1, 54) = 0.001, p = .98, \eta_p^2 = .000$, or violated expectations, $F(1, 54) = 2.68, p = .11, \eta_p^2 = .05$, so we do not discuss these measures further.

would then be an effective tool for defending the gender hierarchy (Rudman, Moss-Racusin, Phelan, & Nauts, 2012). A fourth issue is whether the gender disparity we document generalizes to communal occupations (e.g., among nurses and teachers). Communal occupations confer responsibility for developing relationships with others and ethical traits, such as humility, commitment, and trustworthiness, play an important role in community-building (Haidt, 2007). Therefore, to effectively fulfill communal roles, people of both genders may be held to high ethical standards and punished harshly for ethical violations. Alternatively, women could be punished more harshly even in communal occupations if discriminatory punishment is driven by beliefs that women generally are seductive, manipulative, and evil, and must be deterred from succumbing to their natural unethical proclivities. Ultimately, this is an empirical question—one with the potential to make an important theoretical contribution.

Finally, one important question is whether the gender disparity in punishment for ethical violations can provide a new explanation for why women are under-represented at the top of organizational hierarchies and pay scales. Women might not only ascend more slowly in organizations, but also descend more quickly due to the harsher punishments they incur for ethical violations at work. To understand the net effects of gender for career derailment, however, researchers would need to simultaneously consider how often men and women are found guilty of engaging in unethical behavior alongside the severity of punishments meted out for it. An intensified prescription for women to be ethical could lead women to behave more ethically in practice (or to be perceived as doing so), generating a descriptive stereotype of women as more ethical than men. The descriptive stereotype could produce career benefits for women. For instance, women might be charged with unethical behavior less often. To predict professional outcomes, researchers should attend to both prescriptive and descriptive gender stereotypes.

Reconciling our Findings with Observable Patterns

Before we conclude, we consider how our findings might be reconciled with observable patterns in society, where men seem to be more heavily punished than women given that 93% of prison inmates are male (Federal Bureau of Prisons, 2017). Our research focused on the workplace and did not examine punishments for crimes. The relationship that we uncover may not be generalizable to criminal punishment involving jail sentences for at least three reasons. First, gender differences in punishment for crimes are a function of many factors, including the rate at which women and men commit crimes, the frequency and severity of those crimes, the rate at which offenders are caught, and so on. Women and men could differ along any of these dimensions and then the problem we document might not be noticeable to the naked eye. Second, punishments for crimes may be driven more by race than by gender. African Americans are stereotyped as hostile and dangerous (Bargh, Chen, & Burrows, 1996) and are imprisoned at much higher rates than Caucasians (NAACP, 2017). To understand disparities in punishments for crimes, it is therefore necessary to explore the intersection of race and gender stereotypes. Third, we focus on the retributivist motive for punishment, which holds that people should be punished in accordance with what they deserve. Punishment for violent crimes often reflects the desire to deter future crimes and protect society by removing a threat. Men might receive harsher jail sentences for violent crimes under the logic of deterrence or incapacitation, regardless of what they are perceived to deserve.

Conclusion

Our investigation breaks new ground by describing an important, but under-studied form of gender discrimination: harsher punishments for women than men following ethical violations at work. This disparity emerged in both field and laboratory settings, illustrating its robustness. Our findings represent an important problem for organizations and societies given their interests in the equitable allotment of punishment. We extend knowledge of gender-discrepant outcomes in the workplace by highlighting the fact that gender shapes the allocation of aversive *punishments* for ethical violations.

However, organizations may be able to correct this bias by ensuring that women are equally-represented among the decision-makers responsible for allocating punishments for such transgressions.

REFERENCES

- Albonetti, C. A. 1998. Direct and indirect effects of case complexity, guilty pleas, and offender characteristics on sentencing for offenders convicted of a white-collar offense prior to sentencing guidelines. *Journal of Quantitative Criminology*, 14: 353-378.
- Alicke, M. D. 2000. Culpable control and the psychology of blame. *Psychological Bulletin*, 126: 556-574.
- Amanatullah, E. T., & Morris, M. W. 2010. Negotiating gender roles: Gender differences in assertive negotiating are mediated by women's fear of backlash and attenuated when negotiating on behalf of others. *Journal of Personality and Social Psychology*, 98: 256-267.
- Amanatullah, E. T., & Tinsley, C. H. 2013. Punishing female negotiators for asserting too much...or not enough: Exploring why advocacy moderates backlash against assertive female negotiators. *Organizational Behavior and Human Decision Processes*, 120: 110-122.
- American Bar Association. 2016. Survey on Lawyer Discipline Systems. Retrieved from https://www.americanbar.org/content/dam/aba/administrative/professional_responsibility/2016sol_d_results.authcheckdam.pdf
- Antonio, A. L., Chang, M. J., Hakuta, K., Kenny, D. A., Levin, S., & Milem, J. F. 2004. Effects of racial diversity on complex thinking in college students. *Psychological Science*, 15: 507-510.
- Aquino, K., & Reed, A. II. 2002. The self-importance of moral identity. *Journal of Personality and Social Psychology*, 83: 1423-1440.
- Arvey, R. D., & Ivancevich, J. M. 1980. Punishment in organizations: A review, propositions, and research suggestions. *Academy of Management Review*, 5: 123-132.
- Ayres, I., & Siegelman, P. 1995. Race and gender discrimination in bargaining for a new car. *American Economic Review*, 85: 304-321.
- Bargh, J. A., Chen, M., & Burrows, L. 1996. Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology*, 71: 230-244.
- Belliveau, M. A. 2005. Blind ambition? The effects of social networks and institutional sex composition on the job search outcomes of elite coeducational and women's college graduates. *Organization Science*, 16: 134-150.
- Belliveau, M. A. 2012. Engendering inequity? How social accounts create vs. merely explain unfavorable pay outcomes for women. *Organization Science*, 23: 1154-1174.
- Berdahl, J. L. 2007. The sexual harassment of uppity women. *Journal of Applied Psychology*, 92: 425-437.
- Bowles, H. R., Babcock, L., & Lai, L. 2007. Social incentives for gender differences in the propensity to initiate negotiations: Sometimes it does hurt to ask. *Organizational Behavior and Human Decision Processes*, 103: 84-103.
- Bowles, H. R., & Gelfand, M. 2010. Status and the evaluation of workplace deviance. *Psychological Science*, 2: 49-54.
- Brescoll, V. L. 2011. Who takes the floor and why: Gender, power, and volubility in organizations. *Administrative Science Quarterly*, 56: 622-641.
- Brescoll, V. L., Dawson, E., & Uhlmann, E. L. 2010. Hard won and easily lost: The fragile status of leaders in gender-stereotype-incongruent occupations. *Psychological Science*, 21: 1640-1642.
- Brooks, A. W., Huang, L., Kearney, S. W., & Murray, F. E. 2015. Investors prefer entrepreneurial ventures pitched by attractive men. *PNAS*, 111: 4427-4431.
- Burgess, D., & Borgida, E. 1999. Who women are, who women should be: Descriptive and prescriptive gender stereotyping in sex discrimination. *Psychology, Public Policy, and Law*, 5: 665-692.
- Butterfield, K. D., Treviño, L. K., Wade, K. J., & Ball, G. A. 2005. Organizational punishment from the manager's perspective: An exploratory study. *Journal of Managerial Issues*, 17: 363-382.
- Carlsmith, K. M., Darley, J. M., & Robinson, P. H. 2002. Why do we punish? Deterrence and just deserts as motives for punishment. *Journal of Personality and Social Psychology*, 83: 284-299.
- Castilla, E. J. 2008. Gender, race, and meritocracy in organizational careers. *American Journal of Sociology*, 113: 1479-1526.

- Castilla, E. J., & Benard, S. 2010. The paradox of meritocracy in organizations. *Administrative Science Quarterly*, 55: 543-676.
- Chatman, J. A., & Flynn, F. J. 2005. Full-cycle micro-organizational behavior research. *Organization Science*, 16: 434-447.
- Cohen-Charash, Y., & Spector, P. E. 2001. The role of justice in organizations: A meta-analysis. *Organizational Behavior and Human Decision Processes*, 86: 278-321.
- Cohen, T. R., Gunia, B. C., Kim-Jun, S. Y., & Murnighan, J. K. 2009. Do groups lie more than individuals? Honesty and deception as function of strategic self-interest. *Journal of Experimental Social Psychology*, 45: 1321-1324.
- Cohen, T. R. Panter, A. T., Turan, N., Morse, L., & Kim, Y. 2014. Moral character in the workplace. *Journal of Personality and Social Psychology*, 107: 943-963.
- Cook, A., & Glass, C. 2014. Above the glass ceiling: When are women and racial/ethnic minorities promoted to CEO? *Strategic Management Journal*, 35: 1080-1089.
- Dawson, L. M. 1992. Will feminization change the ethics of the sales profession? *Journal of Personal Selling & Sales Management*, 13: 21-32.
- Desai, S. D., Chugh, D., & Brief, A. P. 2014. The implications of marriage structure for men's workplace attitudes, beliefs, and behaviors toward women. *Administrative Science Quarterly*, 59: 330-365.
- Dreher, G. F., & Cox, T. H. 2000. Labor market mobility and cash compensation: The moderating effects of race and gender. *Academy of Management Journal*, 43: 890-900.
- Duguid, M. M., Lloyd, D. L., & Tolbert, P. S. 2012. The impact of categorical status, numeric representation, and work group prestige on preference for demographically similar others: A value threat approach. *Organization Science*, 23: 386-401.
- Eagly, A. H., Diekmann, A. B., Johannesen-Schmidt, M. C., & Koenig, A. M. 2004. Gender gaps in sociopolitical attitudes: A social psychological analysis. *Journal of Personality and Social Psychology*, 87: 796-816.
- Eagly, A. H., & Mladinic, A. 1989. Gender stereotypes and attitudes toward women and men. *Personality and Social Psychology Bulletin*, 15: 543-558.
- Ely, R. J. 1995. The power in demography: Women's social constructions of gender identity at work. *Academy of Management Journal*, 38, 589-634.
- Federal Bureau of Prisons. 2017. Inmate gender. Retrieved from https://www.bop.gov/about/statistics/statistics_inmate_gender.jsp
- Federation of State Medical Boards. 2016. U.S. medical regulatory trends and actions. Retrieved from <https://www.fsmb.org/globalassets/advocacy/publications/us-medical-regulatory-trends-actions.pdf>
- Fernandez-Mateo, I. 2009. Cumulative gender disadvantage in contract employment. *American Journal of Sociology*, 114: 871-923.
- Fiske, S. T., Cuddy, A. J., Glick, P., & Xu, J. 2002. A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, 82: 878-902.
- Fragale, A. R., Rosen, B., Xu, C., & Merideth, I. 2009. The higher they are, the harder they fall: The effects of transgressor status on observer punishment recommendations and intentionality attributions. *Organizational Behavior and Human Decision Processes*, 108: 53-65.
- Franke, G. R., Crown, D. F., & Spake, D. F. 1997. Gender differences in ethical perceptions of business practices: A social role theory perspective. *Journal of Applied Psychology*, 82: 920-934.
- Glick, P., & Fiske, S. T. 1996. The ambivalent sexism inventory: Differentiating hostile and benevolent sexism. *Journal of Personality and Social Psychology*, 70: 491-512.
- Glick, P., & Fiske, S. T. 2001. An ambivalent alliance: Hostile and benevolent sexism as complementary justifications for gender inequality. *American Psychologist*, 56: 109-118.
- Goldin, C. & Rouse, C. 2000. Orchestrating impartiality: The impact of 'blind' auditions on female musicians." *American Economic Review*, 90: 715-41.

- Goodwin, G. P. 2015. Moral character in person perception. *Current Directions in Psychological Science*, 24: 38-44.
- Goodwin, G. P., Piazza, J., & Rozin, P. 2014. Moral character predominates in person perception and evaluation. *Journal of Personality and Social Psychology*, 106: 148-168.
- Haidt, J. 2007. The new synthesis in moral psychology. *Science*, 316: 998-1002.
- Heilman, M. E. 2012. Gender stereotypes and workplace bias. *Research in Organizational Behavior*, 32: 113-135.
- Iacobucci, D., Posavac, S., Kardes, F. R., Schneider, M. J., & Popovich, D. L. 2015. Toward a more nuanced understanding of the statistical properties of a median split. *Journal of Consumer Psychology*, 25: 652-665.
- Jackall, R. 2010. *Moral mazes: The world of corporate managers*. New York: Oxford University Press.
- Joshi, A. 2014. By whom and when is women's expertise recognized? The interactive effects of gender and education in science and engineering teams. *Administrative Science Quarterly*, 59: 202-239.
- Kanter, R. M. 1977. *Men and women of the corporation*. New York: Basic books.
- Kennedy, J. A., & Anderson, C. 2017. Hierarchical rank and principled dissent: How holding higher rank suppresses objection to unethical practices. *Organizational Behavior and Human Decision Processes*, 139, 30-49.
- Kennedy, J. A., Kray, L. J., & Ku, G. 2017. A social-cognitive approach to understanding gender differences in negotiator ethics: The role of moral identity. *Organizational Behavior and Human Decision Processes*, 138: 28-44.
- Kozlowski, S. W. J., & Klein, K. J. 2000. A multilevel approach to theory and research in organizations: Contextual, temporal, and emergent processes. In K. J. Klein & S. W. J. Kozlowski (Eds.), *Multilevel theory, research and methods in organizations: Foundations, extensions, and new directions* (pp. 3-90). San Francisco, CA: Jossey-Bass.
- Lerner, J. S., & Tetlock, P. E. 1999. Accounting for the effects of accountability. *Psychological Bulletin*, 125: 255-275.
- Longenecker, C. O., Sims, H. P., & Gioia, D. A. 1987. Behind the mask: The politics of employee appraisal. *Academy of Management Executive*, 1, 183-193.
- Martell, R. F., Emrich, C. G., & Robinson-Cox, J. 2012. From bias to exclusion: A multilevel emergent theory of gender segregation in organizations. *Research in Organizational Behavior*, 32: 137-162.
- Mayer, D. M., Nurmohamed, S., Trevino, L. K., Shapiro, D. L., & Schminke, M. 2013. Encouraging employees to report unethical conduct internally: It takes a village. *Organizational Behavior and Human Decision Processes*, 121: 89-103.
- McGinn, K. L., & Milkman, K. L. 2013. Looking up and looking out: Career mobility effects of demographic similarity among professionals. *Organization Science*, 24: 1041-1060.
- Mill, J. S. 1869. *The subjection of women*. London: Longmans, Green, Reader, & Dyer.
- Mooijman, M., van Dijk, W. W., Ellemers, N., & van Dijk, E. 2015. Why leaders punish: A power perspective. *Journal of Personality and Social Psychology*, 109: 75-89.
- Murg, G. E., & Scharman, C. 1982. Employment at will: Do the exceptions overwhelm the rule? *Boston College Law Review*, 2, 329-384.
- NAACP. 2017. Criminal justice fact sheet. Retrieved from <http://www.naacp.org/criminal-justice-fact-sheet/>
- Nadler, J. & McDonnell, M-H. 2012. Moral character, motive, and the psychology of blame. *Cornell Law Review*, 97, 255-304.
- National Business Ethics Survey. 2013. NBES of the U.S. workforce: Key findings. Retrieved from <https://ethics.org/ecihome/research/nbes/nbes-reports/nbes-2013>
- Nordgren, L. & McDonnell, M-H. 2011. The scope-severity paradox: Why doing more harm is judged to be less harmful. *Social Psychological and Personality Science*, 2: 97-102.
- Okimoto, T. G., & Brescoll, V. L. 2010. The price of power: Power seeking and backlash against female politicians. *Personality and Social Psychology Bulletin*, 36: 923-936.

- Peterson, R. S., & Nemeth, C. J. 1996. Focus versus flexibility: Majority and minority influence can both improve performance. *Personality and Social Psychology Bulletin*, 22: 14-23.
- Prentice, D. A., & Carranza, E. 2002. What women and men should be, shouldn't be, are allowed to be, and don't have to be: The contents of prescriptive gender stereotypes. *Psychology of Women Quarterly*, 26: 269-281.
- Rehg, M. T., Miceli, M. P., Near, J. P., & Van Scotter, J. R. 2008. Antecedents and outcomes of retaliation against whistleblowers: Gender differences and power relationships. *Organization Science*, 19: 221-240.
- Ridgeway, C. L. 2001. Gender, status, and leadership. *Journal of Social Issues*, 57: 637-655.
- Rudman, L., & Fairchild, K. 2004. Reactions to counterstereotypic behavior: The role of backlash in cultural stereotype maintenance. *Journal of Personality and Social Psychology*, 87: 157-176.
- Rudman, L. A., Moss-Racusin, C. A., Phelan, J. E., & Nauts, S. 2012. Status incongruity and backlash effects: Defending the gender hierarchy motivates prejudice against female leaders. *Journal of Experimental Social Psychology*, 48: 165-179.
- Rudman, L. A., & Phelan, J. E. 2008. Backlash effects for disconfirming gender stereotypes in organizations. *Research in Organizational Behavior*, 28: 61-79.
- Ryan, M. K., & Haslam, S. A. 2007. The glass cliff: Exploring the dynamics surrounding the appointment of women to precarious leadership positions. *Academy of Management Review*, 32: 549-572.
- Ryan, M. K., Haslam, S. A., Hersby, M. D., & Bongiorno, R. 2011. Think crisis—think female: The glass cliff and contextual variation in the think manager—think male stereotype. *Journal of Applied Psychology*, 96: 470-484.
- Sommers, S. R. 2006. On racial diversity and group decision making: Identifying multiple effects of racial composition on jury deliberations. *Journal of Personality and Social Psychology*, 90: 597-612.
- Stanford Encyclopedia of Philosophy. 2015. Punishment. Retrieved from <https://plato.stanford.edu/entries/punishment/>
- Steffansmeier, D.J., Ulmer, J., & Kramer, J. 1998. The interaction of race, gender, and age in criminal sentencing: The punishment cost of being young, black, and male. *Criminology*, 36(3): 763-98.
- Stinson, P. M., Liederbach, J., Lab, S. P., & Brewer, S. L. 2016. Police integrity lost: A study of law enforcement officers arrested. Technical report, U.S. Department of Justice. Retrieved from <https://www.ncjrs.gov/pdffiles1/nij/grants/249850.pdf>
- Tetlock, P. E. 2002. Social functionalist frameworks for judgment and choice: Intuitive politicians, theologians, and prosecutors. 2002. *Psychological Review*, 109: 451-471.
- Tetlock, P. E., Visser, P. S., Singh, R., Polifroni, M., Scott, A., Elson, S. B., Mazzocco, P., & Rescober, P. 2007. People as intuitive prosecutors: The impact of social-control goals on attributions of responsibility. *Journal of Experimental Social Psychology*, 43: 195-209.
- Trevino, L. K. 1992. The social effects of punishment in organizations: A justice perspective. *Academy of Management Review*, 17: 647-676.
- Uhlmann, E. L., & Cohen, G. L. 2005. Constructed criteria: Redefining merit to justify discrimination. *Psychological Science*, 16: 474-480.
- Vadera, A. K., & Pratt, M. G. 2013. Love, hate, ambivalence, or indifference? A conceptual examination of workplace crimes and organizational identification. *Organization Science*, 24: 172-188.
- Wiltermuth, S. S., & Flynn, F. J. 2013. Power, moral clarity, and punishment in the workplace. *Academy of Management Journal*, 56: 1002-1023.
- Wright, E. O., Baxter, J., & Birkelund, G. E. 1995. The gender gap in workplace authority: A cross-national study. *American Sociological Review*, 60: 407-435.

TABLE 1
Descriptive Statistics and Correlations among Variables in Studies 1 and 2

Variable	N	M	SD	1	2	3	4
Study 1							
1. Female attorney gender	482	0.17	0.37	--			
2. Punishment severity	482	1.89	1.31	.08	--		
3. Total rules violated	482	3.13	2.13	.02	.32***	--	
4. Number of judges on panel	270	5.36	2.02	-.01	-.37	-.07	--
5. Percentage of female judges	270	0.25	0.19	-.11	.16*	.05	.30***
Study 2							
1. Female target gender	57	0.44	0.50	--			
2. Punishment intentions	57			.27*	--		
3. Prescription to be ethical	57			.27*	.24	--	

* $p < .05$. ** $p < .01$. *** $p < .001$.

TABLE 2

Regression Models Predicting the Severity of Punishment for Ethical violations in Study 1

	Ordinal Logistic Regressions of Categorical Punishment			Probit Regressions of Likelihood of Disbarment	
	Model 1	Model 2	Model 3	Model 4	Model 5
<i>Independent Variables</i>					
Female Gender	0.449*	0.892**	1.711***	0.517*	1.429***
	(0.20)	(0.33)	(0.46)	(0.23)	(0.35)
Female Gender x % Female Judges			-3.813*		-4.463**
			(1.75)		(1.49)
<i>Control Variables</i>					
Total Rules Violated	0.291	0.547	0.536	0.826**	0.817**
	(0.23)	(0.34)	(0.34)	(0.25)	(0.24)
Number of Judges on Panel		0.083	0.104	0.093	0.127
		(0.11)	(0.34)	(0.08)	(0.09)
% of Female Judges on Panel		-0.759	-0.495	-0.457	-0.032
		(1.31)	(1.30)	(0.87)	(0.86)
Fixed Effects for Rules Violated	YES	YES	YES	YES	YES
Fixed Effects for State	YES	YES	YES	YES	YES
Constant	^^	^^	^^	-2.319***	-2.680***
				(0.64)	(0.61)
Observations	482	270	270	270	270
Log Pseudolikelihood	-542.58	-285.04	-283.40	-110.48	-107.43
Pseudo R ²	0.17	0.20	0.20	0.28	0.30

Robust standard errors, clustered by state, in parentheses.

^^Because ordinal logistic regression results in multiple intercepts, we have omitted the constant for these models.

* $p < .05$. ** $p < .01$. *** $p < .001$.

TABLE 3
Linear Regression Models Predicting Punishment Severity in Study 2

Variable^a	Model 1		Model 2		Model 3	
	<i>B</i>	<i>SE</i>	<i>B</i>	<i>SE</i>	<i>B</i>	<i>SE</i>
Target's gender ^b	0.46*	.23	0.38	.23	0.26	.23
Strength of the prescription to be ethical			0.30	.23	0.67*	.28
Target gender X Ethical prescription					1.15*	.53
Constant	-.07	.11	-0.08	.11	-0.17	.11
<i>R</i> ²	7%		10%		17%	
<i>df</i>	55		54		53	

^a Variables are mean-centered.

^b Dummy variable coded 0 for male target and 1 for female target.

† $p \leq .10$. * $p \leq .05$. ** $p \leq .01$. *** $p \leq .001$.

FIGURE 1

Predicted Likelihood of Disbarment by Attorney's Gender in Study 1

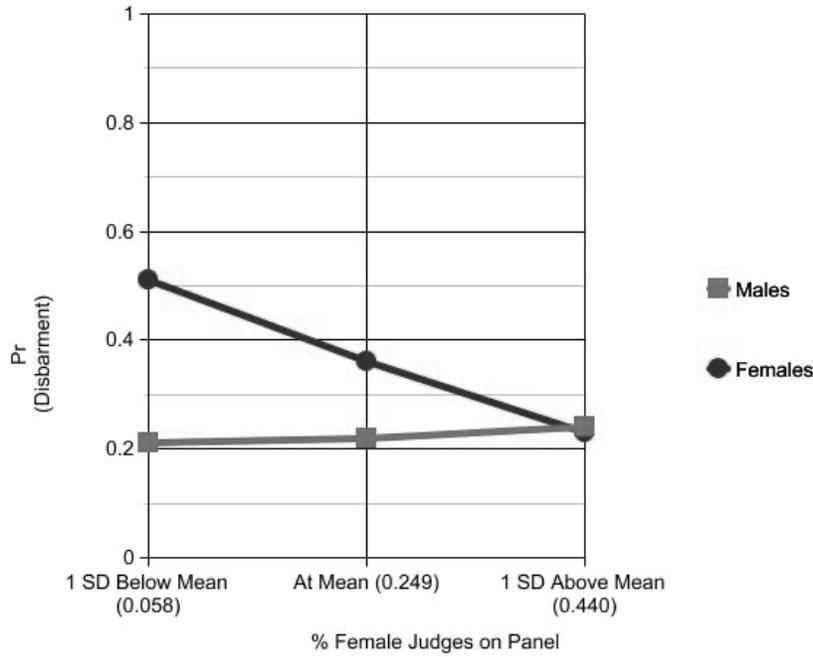
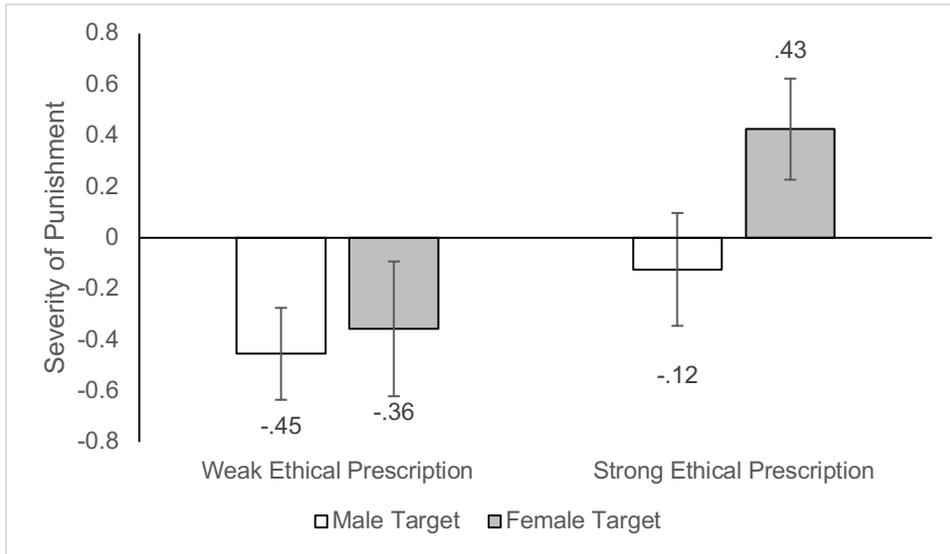


FIGURE 2

Punishment by Target's Gender and Strength of the Prescription to be Ethical in Study 2



Note. Error bars represent +/- 1 standard error of the mean.