

Expertise Fails to Attenuate Gendered Biases in Judicial Decision-Making

Social Psychological and
Personality Science
2019, Vol. 10(2) 227-234
© The Author(s) 2018
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1948550617741181
journals.sagepub.com/home/spp



Andrea L. Miller¹

Abstract

Although the influence of gender ideology on lay decision-making has been established, it is not known to what extent expertise may mitigate gendered biases and improve decision-making quality. In a set of controlled experiments, trial court judges and laypeople evaluated a hypothetical child custody case and a hypothetical employment discrimination case. The role of expertise was tested in two ways: by comparing judges' and laypeople's decision-making and by examining relative differences in expertise among judges. Judges were no less influenced by litigant gender and by their own gender ideology than the lay sample. Judges with greater subject-matter expertise were also no less influenced by gender ideology than other judges. In some cases, expertise was associated with greater, not less, bias. The results of this study suggest that expertise does not attenuate gendered biases in legal decision-making.

Keywords

expertise, bias, decision-making, gender ideology, judges, discrimination

Gender bias plays a significant role in laypeople's decision-making in many domains (e.g., Biernat & Vescio, 2002; Cuddy, Fiske, & Glick, 2004; Lachance-Grzela & Bouchard, 2010; Rudman & Phelan, 2010). Individuals who support traditional gender roles to a greater degree generally believe that women have a natural fit with the domestic sphere and men have a natural fit with career-oriented roles (see Miller & Borgida, 2016). Traditional gender ideologies lead to the perception that women with caregiving responsibilities are bad employees (e.g., S. Correll, Benard, & Paik, 2007; Crosby, Williams, & Biernat, 2004; Fuegen, Biernat, Haines, & Deaux, 2004; Jordan & Zitek, 2012), that women who work are bad mothers (e.g., Okimoto & Heilman, 2012), and that men with caregiving responsibilities are bad employees who violate norms of masculinity (e.g., Berdahl & Moon, 2013; Rudman & Mescher, 2013). Individuals' gender ideology has even been shown to influence legal decision-making. Both real juries and mock juries have been shown to be influenced by gender in decisions such as compensatory damages for lost income (Girvan & Marek, 2016), labor arbitration rulings (Girvan, Deason, & Borgida, 2015), and damage awards in wrongful death suits (Goodman, Loftus, Miller, & Greene, 1991).

It is not known to what extent expertise may mitigate these gender disparities and improve decision-making quality (see Kim & Tidwell, 2014). Judges have substantial subject-matter and decision-making expertise that laypeople lack, and we may expect that this expertise buffers them against the gender biases that characterize laypeople's decisions. Bench trials, in which a judge performs the fact-finding role in place of a

jury, comprise a substantial proportion of civil trials; estimates range between 30% of all civil trials (Langton & Cohen, 2008) and much higher (Hannaford-Agor, Graves, & Miller, 2015). It is important for psychologists to understand, therefore, whether judges are influenced by their ideas about traditional gender roles to the same extent as laypeople.

There is significant ambiguity within psychology on the role of expertise in decision-making quality (Kahneman & Klein, 2009). Some research suggests that judges' decision-making expertise may help them guard against the effects of cognitive biases. For example, Schmittat and Englich (2016) found that judges who had expertise in a particular domain engaged in less confirmatory information processing than laypeople (see also Rachlinski, Guthrie, & Wistrich, 2011).

In contrast to these findings, however, several research studies suggest that judges' decision-making expertise does not protect them against common cognitive biases (e.g., Rachlinski, Wistrich, & Guthrie, 2015; Guthrie, Rachlinski, & Wistrich, 2001). One study showed that judges fell prey to the anchoring and adjustment heuristic during sentencing; they anchored their sentences on a given sentencing demand even if

¹Department of Psychology, University of Illinois at Urbana-Champaign, Champaign, IL, USA

Corresponding Author:

Andrea L. Miller, Department of Psychology, University of Illinois at Urbana-Champaign, 603 East Daniel Street, Champaign, IL 61820, USA.
Email: amiller0@illinois.edu

the demand came from an irrelevant source, if they were informed that the demand was randomly generated, or if they randomly generated the demand themselves by rolling dice (Englich, Mussweiler, & Strack, 2006). Another study found that expertise failed to protect judges against the camera-perspective effect when evaluating videotaped confessions (Lassiter, Diamond, Schmidt, & Elek, 2007). These findings call into question the extent to which expertise may immunize judges against the effects of gendered biases in their decision-making.

The role of expertise in reducing decision-making biases has important implications for the broader psychological literature on discrimination. Critics of laboratory studies on discrimination often argue that the results may not generalize to professional decision-making, where decision makers have expertise and accountability (see, e.g., Tetlock & Mitchell, 2009). If this were true, then decision-making in the real world would be substantially more equitable than decades of research in social psychology suggest. If, however, expertise is not a buffer against decision-making biases, psychologists must focus on identifying interventions. This study examined for the first time the role of judicial expertise in gendered disparities in legal outcomes by directly comparing lay and judicial decision-makers on the same decision tasks.

Research Overview

In a controlled experiment, this study examined the decision-making processes of both laypeople and sitting trial court judges using two hypothetical cases. After the study was completed with the judicial sample, the study procedure was repeated with a lay sample. The hypotheses for this study are as follows:

Hypothesis 1: *The role of gender ideology in judicial decision-making:* Judges' support for traditional gender ideology will correspond to gender disparities in case outcomes.

Hypothesis 2: *The role of expertise in decision-making quality:* (a) Judicial expertise will attenuate gender disparities in case outcomes, such that the judicial sample exhibits less biased decision-making than the lay sample and (b) judicial expertise will attenuate gender disparities in case outcomes, such that judges with more specialized expertise in a particular area of law will exhibit less biased decision-making than judges with less specialized expertise.

Method

The data for the judicial sample come from the 2017 Judicial Decision-Making Study, conducted by the author as part of a larger project. All dependent variables that were analyzed for this article's target research question are reported below. Data collection for the judicial sample took place during a 2-week window. Every trial court judge in the state where the study took place was contacted by his or her chief judge with a

recruitment e-mail and a link to participate in the study online. It was decided in advance that data collection would stop on a particular date, and this decision rule was followed. For the lay sample, it was decided in advance that data collection would stop after a sample size of 500 was achieved. Data were not analyzed before this time. No individual observations were excluded from the analysis below, and all relevant independent variables are reported.

Participants

Judicial sample. Participants in the judicial sample were 619 sitting trial court judges from a state jurisdiction in the United States. The sample comprised 68.7% of all trial judges in the state. Participants identified as 85.5% White, 5.2% Black, 2.9% Latinx, 1.8% Asian, 1.0% multiracial, 0.2% Native American, and 2.1% Other (1.5% did not identify a race or ethnicity). Participants identified their gender as 68.2% men, 29.7% women, and 0.6% other (1.5% did not identify a gender). The sample was representative of the full statewide population of judges with regard to gender and was moderately representative with regard to geographic region (for an analysis of sample representativeness, see the Supplementary Online Material [SOM]).

Lay sample. Participants in the lay sample were 504 adults from across the United States that were recruited through Mechanical Turk in August 2017. The mean age for the sample was 34.35, and the sample ranged from 18 to 75 years old. Participants identified as 77.8% White, 7.9% Black, 7.1% Asian, 3.6% Latinx, 2.6% multiracial, 0.4% Native American, and 0.4% Other (0.2% did not identify a race or ethnicity). Participants identified their gender as 59.1% men and 40.9% women.

Design

The 2017 Judicial Decision-Making Study employed a factorial design in each hypothetical court case. The child custody case used a 2 (race) \times 2 (income) \times 2 (gender) mixed design. Judges evaluated a hypothetical divorce and child custody case in which the litigants were either Black or White and were either high or low income. The litigants were a male and female spouse, so gender was varied within subjects. Because race and income had no significant effect on the results, all findings presented below collapse across race and income (see SOM for race and income analyses).

The employment discrimination case used a 2 (race) \times 2 (gender) between-subjects design. Judges evaluated a hypothetical civil case in which the plaintiff was either Black or White and either a man or a woman. Because race had no significant effect on the results, all findings presented below collapse across race (see SOM for race analyses).

Statistical power. Previous research does not provide a clear basis by which to predict a particular effect size for this study. A priori power analysis suggested that in order to detect a small

Table 1. Bivariate Pearson Correlations Among All Predictor and Outcome Variables in the Judicial Sample.

	Variable								
	1	2	3	4	5	6	7	8	9
1. Gender ideology	1	-.11*	.08*	.07	-.06	-.15*	-.03	-.02	.01
2. Specialization in criminal law		1	-.73*	-.44*	-.02	.06	-.03	.13*	.04
3. Specialization in civil law			1	-.13*	-.06	.01	-.01	-.06	-.01
4. Specialization in family law				1	-.02	-.14*	-.01	-.02	-.08
5. Legal custody					1	.14*	-.02	-.01	-.03
6. Physical custody						1	-.03	-.05	.04
7. Perceived case merit							1	-.44*	.41*
8. Summary judgment								1	-.21*
9. Final case ruling									1

*p < .05.

effect size in a *t* test (Cohen’s *d* = .20) at .80 power, a sample size of 310 per group would be required (Soper, 2017b). The *t* tests in this study compare the judicial sample (*n* = 619) to the lay sample (*n* = 504), so the group sizes far exceed this suggested minimum.

A priori power analysis suggested that in order to detect a small effect size in multiple regression (Cohen’s *f*² = .02) at .80 power with an interaction term (three total predictors), a sample size of 543 would be required (Soper, 2017a); in order to detect a medium effect size (Cohen’s *f*² = .15) under these conditions, a sample size of 76 would be required. This study included regression analyses with an interaction term among the judicial sample alone (*n* = 619) and among the combined judicial and lay samples (*N* = 1,123).

Predictor Variables

Gender ideology. Gender ideology was measured using the Separate Spheres Ideology (SSI) Scale (Miller & Borgida, 2016), a 15-item measure of support for traditional social roles for men and women (e.g., “It’s natural for a woman to be fulfilled by taking care of her children, but most men feel better when they have a good career, too”; “Most kids are better off if their dad is the primary provider for the whole family”). This measure predicts a wide variety of attitudes and behaviors reflecting traditional gender norms and the belief that men and women fit in different spheres of society. Participants responded to each item on a 7-point Likert-type scale ranging from *strongly disagree* to *strongly agree* (judicial sample: *M* = 2.43, standard deviation [*SD*] = 0.84, α = .85; lay sample: *M* = 2.94, *SD* = 1.20, α = .92; see Tables 1 and 2 for correlations among all predictor and outcome variables).

Expertise. Expertise was examined in two ways. First, the decision-making of judges was compared directly to the decision-making of laypeople. Second, relative differences in expertise among judges was measured by their specialization in different areas of law. Participants indicated on binary measures whether they had practiced as an attorney or presided as a judge over criminal cases, civil cases, and family cases in the

Table 2. Bivariate Pearson Correlations Among All Predictor and Outcome Variables in the Lay Sample.

	Variable					
	1	2	3	4	5	6
1. Gender ideology	1	-.06	-.02	-.25*	.25*	-.36*
2. Legal custody		1	.32*	-.03	-.04	.09*
3. Physical custody			1	-.01	.09*	-.01
4. Perceived case merit				1	-.37*	.45*
5. Summary judgment					1	-.47*
6. Final case ruling						1

*p < .05.

last 10 years. Participants then indicated what percentage of their current caseload was comprised of each area of law (judges in this jurisdiction are assigned caseloads in a wide variety of ways—case assignments are sometimes, but not always, random); these percentages served as a measure of specialization (criminal: *M* = 45.08, *SD* = 42.20; civil: *M* = 38.60, *SD* = 41.47; and family: *M* = 19.48, *SD* = 32.07).

Outcome Variables: Gendered Disparities in Case Outcomes

Legal custody in a divorce case. Participants evaluated a divorce case in which two spouses were contesting custody of their children. The measure of legal custody (How will you allocate significant decision-making responsibility?) was a 3-point item in which participants chose between *custody for the woman* (–1), *custody for both spouses* (0), or *custody for the man* (1; judicial sample: *M* = –0.04, *SD* = 0.21; lay sample: *M* = –0.06, *SD* = 0.37).

Physical custody in a divorce case. The measure of physical custody (How will you allocate parenting time?) was an 11-point item ranging from *full physical custody for the woman* (–5) to *shared equal custody* (0) to *full physical custody for the man* (5; judicial sample: *M* = –0.46, *SD* = 0.84; lay sample: *M* = –0.15, *SD* = 1.32).

Perceived merit of a discrimination case. Participants also evaluated an employment discrimination case. The measure of perceived case merit (Based on this information, what is your overall impression of the merit of Jones' case?) was a 6-point item ranging from *no merit* to *very high merit* (judicial sample: $M = 3.02$, $SD = 1.01$; lay sample: $M = 3.17$, $SD = 1.26$).

Summary judgment in a discrimination case. Participants determined whether to grant the defendant's motion for summary judgment, which would dismiss the plaintiff's case. The summary judgment ruling (Based on this information, what is your ruling?) was measured using a binary item in which participants *granted* (1) or *denied* (0) summary judgment (judicial sample: $M = 0.13$, $SD = 0.33$; lay sample: $M = 0.34$, $SD = 0.48$).

Final ruling in the discrimination case. Finally, participants provided a final verdict on whether the defendant had discrimination against the plaintiff. The final ruling (What is your final ruling on the case?) was measured using a binary item in which participants *ruled for* (1) or *against* (0) the plaintiff (judicial sample: $M = 0.66$, $SD = 0.47$; lay sample: $M = 0.54$, $SD = 0.50$).

Procedure

Judicial sample. Participants first completed the measures of legal expertise. Next, participants evaluated several hypothetical cases, including the two cases examined in this study. Participants only evaluated cases in areas of law in which they had experience in the last 10 years; therefore, 372 judges evaluated the child custody case, and 514 judges evaluated the employment discrimination case.

The first case was a child custody dispute between two divorcing spouses. The child custody case (see SOM for full text) was written alongside several legal scholars and judges to make it as realistic as possible. Race and gender were manipulated using the litigants' names, which were chosen from previous research identifying names that are perceived as belonging to people of certain races and genders (Bertrand & Mullainathan, 2004).

The second case was an employment discrimination case in which a plaintiff sued his or her employer for family responsibilities discrimination, a form of sex discrimination that penalizes employees for their real or perceived family caregiving responsibilities (see Miller, 2014). The text of the case (see SOM) was identical across conditions except for the race and gender of the plaintiff, which were again manipulated using names and gendered pronouns (see Table 3 for descriptive statistics by gender condition).

After working through the cases, participants completed the SSI Scale. Finally, participants provided demographic information.

Lay sample. The lay sample completed the same procedure as described above, with two modifications. First, instead of the measure of specialization in different areas of law, participants indicated whether they had any experience with legal decision-making. Filtering participants with legal experiences out of the

Table 3. Descriptive Statistics for the Discrimination Case Outcomes by Gender Condition.

Outcome Variable	Minimum	Maximum	Male Plaintiff		Female Plaintiff	
			M	SD	M	SD
Perceived case merit	1	6	3.02	1.13	3.17	1.15
Summary judgment	0	1	0.25	0.43	0.22	0.42
Final case ruling	0	1	0.56	0.50	0.64	0.48

Note. SD = standard deviation.

analysis did not change any of the results (see SOM). Therefore, the results presented below include all participants.

Second, the text of the hypothetical cases was identical to those that the judicial sample completed, except that extra instructions were added to the questions (see SOM). These extra instructions provided basic definitions and decision rules that judges would know, but laypeople would not. For example, before asking lay participants to rule on the discrimination case, the instructions explained the burden of proof and the evidentiary standard.

Results: Domestic Sphere

Hypothesis 1: The Role of Gender Ideology in Child Custody Case Outcomes

In order to examine gendered biases in judicial decision-making in the domestic sphere, I first examined the role of judges' personal gender ideology in their allocation of legal and physical custody in the divorce case.

Legal custody. In a linear regression, support for the separate spheres ideology did not play a significant role in the allocation of legal custody ($B = 0.01$, standard error [SE] = .01, 95% confidence interval [CI] = [-0.01, 0.04], $\beta = .06$, $p = .280$). The model may have suffered from a restriction of range, as only 17 participants chose to allocate legal custody to a sole spouse (16 of these allocated legal custody to the female litigant).

Physical custody. In contrast, support for the SSI played a significant role in physical custody ($B = 0.15$, $SE = .05$, 95% CI [0.05, 0.25], $\beta = .15$, $p = .004$). The more participants supported traditional gender roles for men and women, the more parenting time they allocated to the female litigant at the expense of the male litigant. This is initial evidence that despite their substantial expertise, judges are not immune to the effects of gendered biases in their decisions.

Hypothesis 2a: The Role of Expertise—Comparing Judges and Laypeople on Custody Outcomes

I next examined the effect of expertise in mitigating gendered biases by comparing the extent to which the judicial sample and the lay sample were influenced by litigant gender.

Legal custody. An independent-samples *t* test revealed that the judicial sample ($M = 0.04$, $SD = 0.21$) and lay sample ($M = 0.06$, $SD = 0.37$) did not differ significantly in their allocation of legal custody to the male and female litigants, 95% CI $[-0.06, 0.02]$, $t(873) = 1.00$, $p = .319$, Cohen's $d = .07$. Thus, the level of expertise that judges possess did not reduce the influence of litigant gender on this decision relative to laypeople.

Physical custody. In contrast, an independent-samples *t* test revealed that the judicial sample ($M = 0.46$, $SD = 0.84$) and lay sample ($M = 0.15$, $SD = 1.32$) differed significantly in their allocation of physical custody, 95% CI $[0.15, 0.46]$, $t(873) = 3.88$, $p < .001$, Cohen's $d = .28$. Post hoc analysis revealed that the judicial sample was influenced *more*, not less, by litigant gender on this task. Whereas the judges allocated approximately an extra half day of parenting time to the female litigant, laypeople only allocated 0.15 of an extra day of parenting time to her.

Hypothesis 2b: The Role of Expertise—Does Specialization Among Judges Influence Custody Outcomes?

I next examined the role of expertise by testing the effect of specialization on judges' decision-making. If expertise improves decision-making quality, it should reduce the influence of gender ideology on case outcomes.

Legal custody. A linear regression revealed that the interaction between specialization in family law and SSI scores was not a significant predictor of legal custody ($B = 0.00$, $SE = .00$, 95% CI $[0.00, 0.00]$, $\beta = .06$, $p = .736$).

Physical custody. The interaction between specialization in family law and SSI scores was likewise not a significant predictor of physical custody ($B = 0.00$, $SE = .00$, 95% CI $[0.00, 0.01]$, $\beta = .21$, $p = .228$). In other words, specialization failed to reduce the influence of judges' gender ideology on their child custody decisions.

Results: Employment Sphere

Hypothesis 1: The Role of Gender Ideology in Employment Discrimination Case Outcomes

I first examined the role of gender ideology in judges' decision-making.

Perceived case merit. In a linear regression, the interaction between judges' SSI scores and plaintiff gender was a significant predictor of perceived merit ($B = -0.29$, $SE = .10$, 95% CI $[-0.49, -0.09]$, $\beta = -.39$, $p = .005$). Post hoc simple slopes analysis revealed that at 1 *SD* above the mean on traditional gender ideology, plaintiff gender had a significant effect on participants' perceived merit of the case ($B = 0.29$,

$SE = .06$, 95% CI $[0.16, 0.41]$, $\beta = .29$, $p < .001$). In contrast, at 1 *SD* below the mean on traditional gender ideology, plaintiff gender had no effect on participants' perceived merit of the case ($B = 0.04$, $SE = .06$, 95% CI $[-0.08, 0.16]$, $\beta = .04$, $p = .470$). These slopes were significantly different from each other, $t(1,006) = 2.95$, $p = .003$. Thus, the more strongly judges supported traditional social roles for men and women, the more their initial perceptions of case merit were influenced by plaintiff gender.

Summary judgment. The interaction between participants' SSI scores and plaintiff gender was also a significant predictor of judges' rulings on the motion for summary judgment, $B = 1.31$, $SE = .34$, $\text{Exp}(B) = 3.72$, 95% CI for $\text{Exp}(B)$ $[1.90, 7.29]$, Wald = 14.62, $p < .001$. Post hoc simple slopes analysis revealed that at 1 *SD* above the mean on traditional gender ideology, plaintiff gender had a significant effect on the likelihood of summary judgment, $B = 2.11$, $SE = .46$, $\text{Exp}(B) = 8.22$, 95% CI for $\text{Exp}(B)$ $[3.33, 20.29]$, Wald = 20.89, $p < .001$. In contrast, at 1 *SD* below the mean on traditional gender ideology, plaintiff gender had no effect on the likelihood of summary judgment, $B = -0.09$, $SE = .38$, $\text{Exp}(B) = 0.92$, 95% CI for $\text{Exp}(B)$ $[0.43, 1.95]$, Wald = 0.05, $p = .819$. These slopes were significantly different from each other, $t(1,006) = 3.69$, $p < .001$. Thus, the more strongly judges supported traditional social roles for men and women, the more their summary judgment rulings were influenced by plaintiff gender.

Final discrimination ruling. The interaction between participants' SSI scores and plaintiff gender was a significant predictor of judges' final ruling in the case, $B = -0.46$, $SE = .23$, $\text{Exp}(B) = 0.63$, 95% CI for $\text{Exp}(B)$ $[0.40, 0.98]$, Wald = 4.13, $p = .042$. Post hoc simple slopes analysis revealed that at 1 *SD* above the mean on traditional gender ideology, plaintiff gender had a significant effect on the likelihood of ruling for the plaintiff, $B = -1.18$, $SE = .28$, $\text{Exp}(B) = 0.31$, 95% CI for $\text{Exp}(B)$ $[0.18, 0.53]$, Wald = 17.88, $p < .001$. In contrast, at 1 *SD* below the mean on traditional gender ideology, plaintiff gender had no effect on the likelihood of ruling for the plaintiff, $B = -0.40$, $SE = .27$, $\text{Exp}(B) = 0.67$, 95% CI for $\text{Exp}(B)$ $[0.40, 1.12]$, Wald = 2.31, $p = .129$. These slopes were significantly different from each other, $t(1,006) = 2.01$, $p = .045$. Thus, the more strongly judges supported traditional social roles for men and women, the more their final rulings on the discrimination case were influenced by plaintiff gender.

Hypothesis 2a: The Role of Expertise—Comparing Judges and Laypeople on Discrimination Case Outcomes

I next compared the extent to which the judicial sample and the lay sample were influenced by plaintiff gender.

Perceived case merit. The interaction between sample and plaintiff gender was a significant predictor of the perceived merit of the plaintiff's initial case ($B = 0.34$, $SE = .14$, 95% CI $[0.06, 0.63]$, $\beta = .13$, $p = .016$). Post hoc analysis revealed that

plaintiff gender was a significant predictor of perceived merit in the judicial sample ($B = 0.32$, $SE = .09$, 95% CI [0.15, 0.50], $\beta = .16$, $p < .001$) but not in the lay sample ($B = -0.02$, $SE = .11$, 95% CI [-0.24, 0.20], $\beta = -.01$, $p = .851$); the slopes for these samples were significantly different, $t(1,005) = 2.39$, $p = .017$. Thus, plaintiff gender had a larger influence on judges' perceptions of case merit than on laypeople's.

Summary judgment. The interaction between sample and plaintiff gender was also a significant predictor of the summary judgment ruling, $B = -1.09$, $SE = .34$, $\text{Exp}(B) = 0.34$, 95% CI for $\text{Exp}(B)$ [0.17, 0.66], Wald = 10.19, $p = .001$. Plaintiff gender was a significant predictor of the probability of summary judgment in the judicial sample, $B = -0.91$, $SE = .29$, $\text{Exp}(B) = 0.40$, 95% CI for $\text{Exp}(B)$ [0.23, 0.71], Wald = 10.14, $p = .001$ but not in the lay sample, $B = 0.18$, $SE = .19$, $\text{Exp}(B) = 1.20$, 95% CI for $\text{Exp}(B)$ [0.83, 1.73], Wald = 0.94, $p = .333$; the slopes for these samples were significantly different, $t(1,005) = 3.14$, $p = .002$. Thus, plaintiff gender had a larger influence on judges' summary judgment rulings than on laypeople's.

Final discrimination ruling. The interaction between sample and plaintiff gender was a significant predictor of the final ruling on discrimination, $B = 0.76$, $SE = .26$, $\text{Exp}(B) = 2.14$, 95% CI for $\text{Exp}(B)$ [1.28, 3.58], Wald = 8.34, $p = .004$. Plaintiff gender was a significant predictor of the final ruling in the judicial sample, $B = 0.77$, $SE = .19$, $\text{Exp}(B) = 2.15$, 95% CI for $\text{Exp}(B)$ [1.48, 3.14], Wald = 15.83, $p < .001$, but not in the lay sample, $B = 0.01$, $SE = .18$, $\text{Exp}(B) = 1.00$, 95% CI for $\text{Exp}(B)$ [0.71, 1.43], Wald = 0.00, $p = .967$; the slopes for these samples were significantly different, $t(1,005) = 2.90$, $p = .004$. Thus, plaintiff gender had a larger influence on judges' case rulings than on laypeople's. Together, these findings suggest that expertise does not reduce the effect of plaintiff gender on decision-making; plaintiff gender played a *larger* role in case outcomes in the judicial sample than in the lay sample.

Hypothesis 2b: The Role of Expertise—Does Specialization Among Judges Influence Discrimination Case Outcomes?

I next examined effect of specialization on judges' decision-making.

Perceived case merit. A linear regression revealed that the interaction between specialization in civil law and SSI scores was not a significant predictor of perceived case merit ($B = 0.00$, $SE = .00$, 95% CI [0.00, 0.00], $\beta = -.04$, $p = .817$).

Summary judgment. The interaction between specialization in civil law and SSI scores was not a significant predictor of the likelihood of summary judgment, $B = 0.00$, $SE = .00$, $\text{Exp}(B) = 1.00$, 95% CI for $\text{Exp}(B)$ [0.99, 1.01], Wald = 0.00, $p = .997$.

Final discrimination ruling. The interaction between specialization in civil law and SSI scores was not a significant predictor of the likelihood of ruling for the plaintiff, $B = 0.00$, $SE = .00$, $\text{Exp}(B) = 1.00$, 95% CI for $\text{Exp}(B)$ [1.00, 1.01], Wald = 0.58, $p = .446$. Together, these results suggest that specialized expertise does not reduce the influence of gender ideology when judges are deciding cases in which gender roles are at play.

General Discussion

Although the influence of gender ideology on lay decision-making has been established, it is not known to what extent expertise may improve decision-making quality. This study examined for the first time the role of judicial expertise in gendered disparities in legal outcomes. The results of this study supported the first hypothesis by showing that judges' decision-making was substantially influenced by gender ideology. Judges' support for traditional gender roles for men and women predicted gender disparities in both a child custody case and an employment discrimination case.

The role of expertise in mitigating these biases was tested in two ways: by comparing the decisions of judges directly to a lay sample and by examining relative differences in expertise among the judges. On both of these metrics, expertise failed to attenuate the effects of gender ideology on judicial decision-making. In cases where expertise made a significant difference, it was associated with *more* gender bias not less. These findings suggest that the expertise that judges possess does not buffer them against the biasing influence of gender ideology. These findings also raise the possibility that expertise may open the door for greater bias in some cases.

Several directions for future research flow from these findings. First, these findings raise the possibility that expertise not only fails to attenuate gender biases in judicial decision-making but may actually exacerbate them. The correlational nature of the results presented here means that it is not possible to infer a causal relationship. Future research should examine the possibility of a causal relationship in greater depth, perhaps by manipulating expertise experimentally or measuring the acquisition of expertise longitudinally. Research in political psychology suggests that we may have reason to expect a positive relationship between expertise and bias. For example, individuals with the most political sophistication and expertise engage in information processing that is characterized by the greatest confirmation and disconfirmation biases and, consequently, endorse the most polarized political attitudes (Taber & Lodge, 2006). This finding raises the possibility that expertise in some domains may give individuals the cognitive tools they need to engage in more biased information processing, rather than less. It is also possible that expertise leads individuals to adopt more heuristic and less systematic information processing, which could open the door for ideology and biases to influence the decision outcome.

Second, future research should examine different types of expertise. For example, judges have substantive expertise in the

content of the law as well as procedural expertise in the task of making complex decisions about cases. It may be the case that different types of expertise play different roles in the reduction of bias in decision-making. Researchers should tease apart these various types of expertise and test their effects separately.

The results of this study have significant implications for psychological research on stereotyping and discrimination. Lively debates about the role of expertise in reducing discrimination exist in research on organizations (e.g., Jost et al., 2009), police work (e.g., J. Correll et al., 2007), and other domains. Generally speaking, expertise has not emerged as a silver bullet for combatting discrimination in these other domains. For example, research on police shooting decisions suggests that although police officers sometimes perform better than the general population on shooter tasks, they still exhibit racial biases in both the accuracy and speed of the decision to shoot (J. Correll et al., 2007; see also Eberhardt, Goff, Purdie, & Davies, 2004). A common critique of laboratory studies showing discrimination in professional settings is that the results may not generalize to the real world, where decision makers have expertise and accountability. The results presented here suggest, in contrast, that expertise is not the panacea for reducing bias in professional settings that we might hope it to be.

Acknowledgments

The author thanks the American Bar Foundation, Bob Nelson, Shari Diamond, Janice Nadler, and Elizabeth Mertz for valuable support and feedback at various stages of the project. The author also thanks each of the judges who participated in the study despite the personal and professional risks.

Declaration of Conflicting Interests

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: The author regularly receives consulting and speaking fees from state court entities. The author does not feel that these professional relationships influenced the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The Judicial Decision-Making Study was made possible by the financial and administrative support of the state supreme court in the state where the study was conducted. The court played no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the article; or in the decision to submit the article for publication.

Supplemental Material

The supplemental material is available in the online version of the article. All research materials and the lay sample data can also be accessed at: <https://doi.org/10.6084/m9.figshare.5306611>

References

- Berdahl, J., & Moon, S. (2013). Workplace mistreatment of middle class workers based on sex, parenthood, and caregiving. *Journal of Social Issues, 69*, 341–366. doi:10.1111/josi.12018
- Bertrand, M., & Mullainathan, S. (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *The American Economic Review, 94*, 991–1013.
- Biernat, M., & Vescio, T. K. (2002). She swings, she hits, she's great, she's benched: Implications of gender-based shifting standards for judgment and behavior. *Personality and Social Psychology Bulletin, 28*, 66–77. doi:10.1177/0146167202281006
- Correll, J., Park, B., Judd, C., Wittenbrink, B., Sadler, M., & Keesee, T. (2007). Across the thin blue line: Police officers and racial bias in the decision to shoot. *Journal of Personality and Social Psychology, 92*, 1006–1023. doi:10.1037/0022-3514.92.6.1006
- Correll, S., Benard, S., & Paik, I. (2007). Getting a job: Is there a motherhood penalty? *American Journal of Sociology, 112*, 1297–1338. doi:10.1086/511799
- Crosby, F., Williams, J., & Biernat, M. (2004). The maternal wall. *Journal of Social Issues, 60*, 675–682. doi:10.1111/j.0022-4537.2004.00379.x
- Cuddy, A., Fiske, S., & Glick, P. (2004). When professionals become mothers, warmth doesn't cut the ice. *Journal of Social Issues, 60*, 701–718. doi:10.1111/j.0022-4537.2004.00381.x
- Eberhardt, J., Goff, P., Purdie, V., & Davies, P. (2004). Seeing black: Race, crime, and visual processing. *Journal of Personality and Social Psychology, 87*, 876–893. doi:10.1037/0022-3514.87.6.876
- Englich, B., Mussweiler, T., & Strack, F. (2006). Playing dice with criminal sentences: The influence of irrelevant anchors on experts' judicial decision making. *Personality and Social Psychology Bulletin, 32*, 188–200. doi:10.1177/0146167205282152
- Fuegen, K., Biernat, M., Haines, E., & Deaux, K. (2004). Mothers and fathers in the workplace: How gender and parental status influence judgments of job-related competence. *Journal of Social Issues, 60*, 737–754. doi:10.1111/j.0022-4537.2004.00383.x
- Girvan, E., Deason, G., & Borgida, E. (2015). The generalizability of gender bias: Testing the effects of contextual, explicit, and implicit sexism on labor arbitration decisions. *Law and Human Behavior, 39*, 525–537. doi:10.1037/lhb0000139
- Girvan, E., & Marek, H. (2016). Psychological and structural bias in civil jury awards. *Journal of Aggression, Conflict and Peace Research, 8*, 247–257. doi:10.1108/JACPR-09-2015-0190
- Goodman, J., Loftus, E., Miller, M., & Greene, E. (1991). Money, sex, and death: Gender bias in wrongful death damage awards. *Law & Society Review, 25*, 263–286. doi:10.2307/3053799
- Guthrie, C., Rachlinski, J., & Wistrich, A. (2001). Inside the judicial mind. *Cornell Law Review, 86*, 777–830.
- Hannaford-Agor, P., Graves, S., & Miller, S. (2015). *The landscape of civil litigation in state courts*. National Center for State Courts Civil Justice Initiative. Retrieved from <https://www.ncsc.org/~media/Files/PDF/Research/CivilJusticeReport-2015.ashx>

- Jordan, A., & Zitek, E. (2012). Marital status bias in perceptions of employees. *Basic and Applied Social Psychology, 34*, 474–481. doi:10.1080/01973533.2012.711687
- Jost, J., Rudman, L., Blair, I., Carney, D., Dasgupta, N., Glaser, J., & Hardin, C. (2009). The existence of implicit bias is beyond reasonable doubt: A refutation of ideological and methodological objections and executive summary of ten studies that no manager should ignore. *Research in Organizational Behavior, 29*, 39–69. doi:10.1016/j.riob.2009.10.001
- Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise: A failure to disagree. *American Psychologist, 64*, 515–526. doi:10.1037/a0016755
- Kim, A., & Tidwell, N. (2014). Examining the impact of sexism on evaluations of social scientific evidence in discrimination litigation. *Law and Human Behavior, 38*, 520–530. doi:10.1037/lhb0000103
- Lachance-Grzela, M., & Bouchard, G. (2010). Why do women do the lion's share of housework? A decade of research. *Sex Roles, 63*, 767–780. doi:10.1007/s11199-010-9797-z
- Langton, L., & Cohen, T. (2008). Civil bench and jury trials in state courts, 2005. Bureau of Justice Statistics Special Report. Retrieved from <https://www.bjs.gov/content/pub/pdf/cbjtsc05.pdf>
- Lassiter, G., Diamond, S., Schmidt, H., & Elek, J. (2007). Evaluating videotaped confessions: Expertise provides no defense against the camera-perspective effect. *Psychological Science, 18*, 224–226. doi:10.1111/j.1467-9280.2007.01879.x
- Miller, A. (2014). The separate spheres ideology: An improved empirical and litigation approach to family responsibilities discrimination. *Minnesota Law Review, 99*, 343–379.
- Miller, A., & Borgida, E. (2016). The separate spheres model of gendered inequality. *PLOS One, 11*. doi:10.1371/journal.pone.0147315
- Okimoto, T. G., & Heilman, M. E. (2012). The “bad parent” assumption: How gender stereotypes affect reactions to working mothers. *Journal of Social Issues, 68*, 704–724. doi:10.1111/j.1540-4560.2012.01772.x
- Rachlinski, J., Guthrie, C., & Wistrich, A. (2011). Probable cause, probability, and hindsight. *Journal of Empirical Legal Studies, 8*, 72–98. doi:10.1111/j.1740-1461.2011.01230.x
- Rachlinski, J., Wistrich, A., & Guthrie, C. (2015). Can judges make reliable numeric judgments? Distorted damages and skewed sentences. *Indiana Law Journal, 90*, 695–739.
- Rudman, L., & Mescher, K. (2013). Penalizing men who request a family leave: Is flexibility stigma a femininity stigma? *Journal of Social Issues, 69*, 322–340. doi:10.1111/josi.12017
- Rudman, L., & Phelan, J. (2010). The effect of priming gender roles on women's implicit gender beliefs and career aspirations. *Social Psychology, 41*, 192–202.
- Schmittat, S., & Englich, B. (2016). If you judge, investigate! Responsibility reduces confirmatory information processing in legal experts. *Psychology, Public Policy, and Law, 22*, 386–400. doi:10.1037/law0000097
- Soper, D. (2017a). *A-priori sample size calculator for multiple regression* [Software]. Retrieved from <http://www.danielsoper.com/statcalc>
- Soper, D. (2017b). *A-priori sample size calculator for student t-tests* [Software]. Retrieved from <http://www.danielsoper.com/statcalc>
- Taber, C., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science, 50*, 755–769. doi:10.1111/j.1540-5907.2006.00214.x
- Tetlock, P., & Mitchell, G. (2009). Implicit bias and accountability systems: What must organizations do to prevent discrimination? *Research in Organizational Behavior, 29*, 3–38. doi:10.1016/j.riob.2009.10.002

Author Biography

Andrea L. Miller is a social psychologist at the University of Illinois. Her research examines the social-psychological foundations of prejudice, discrimination, and inequality, particularly as they interact with the law.

Handling Editor: Kate Ratliff