Ascriptive Characteristics and Perceptions of Impropriety in the Rule of Law: Race, Gender, and Public Assessments of Whether Judges Can Be Impartial

ONO, Yoshikuni
RIETI

ZILIS, Michael A.
University of Kentucky

Ascriptive Characteristics and Perceptions of Impropriety in the Rule of Law: Race, Gender, and Public Assessments of Whether Judges Can Be Impartial[1]

ONO, Yoshikuni

Tohoku University & Research Institute of Economy, Trade and Industry

ZILIS, Michael A.

University of Kentucky

Abstract

Perceptions of procedural fairness influence the legitimacy of the law, and because procedures are mutable, reforming them can buttress support for the rule of law. Yet legal authorities have recently faced a distinct challenge: accusations of impropriety based on their ascriptive characteristics (e.g., gender, ethnicity). We study the effect of these traits in the context of the U.S. legal system, focusing on the conditions under which citizens perceive female and minority judges as exhibiting impropriety, and how this compares with perceptions of their white and male counterparts. We find that Americans use a judge's race and gender to make inferences about which groups the judge favors, whether she is inherently biased, and whether she should recuse. Notably, we find drastically different evaluations of female and Hispanic judges among the political right and left.

Support for the rule of law plays a pivotal role in the functioning and health of democratic systems. A key ingredient shaping obedience and compliance concerns whether the public perceives legal authorities to follow fair and predictable procedures (Tyler 1990; Tyler and Huo 2002; Tyler 2003). Yet procedural fairness is not the only ingredient that influences confidence in the rule of law. In an increasingly polarized environment, legal authorities have come under attack for their *ascriptive* characteristics as well. To take one example, elites have suggested that female and minority judges display prejudice in certain cases. In her confirmation process, Justice Sonia Sotomayor got in some trouble when she said that hoped "a wise Latina woman with the richness of her experiences would more often than not reach a better conclusion than a white male who hasn't lived that life." Sotomayor sought to emphasize the positive aspects of diversity, including the idea that her personal traits may bring different perspectives to the law, but she was criticized from multiple Senators who were concerned about potential bias stemming from those attributes. Similarly, in 2016, Republican presidential nominee Donald Trump made a sustained effort to taint public faith in a federal judge, arguing that the judge could not rule fairly because of his ethnic background.[1]

These accusations exemplify a strategic approach used increasingly in democratic societies to undermine the rule of law, a fact made more concerning because it focuses on ascriptive characteristics. Existing scholarship, however, provides insufficient explanation about how this approach is received by citizens. If the public is receptive to evaluating the legal system based on

---

[1] At one campaign rally, Trump mentioned that the judge, presiding over a class-action lawsuit against Trump University, "happens to be, we believe, Mexican." See Johnson, Jenna, and Philip Rucker. 2016. "In San Diego, Trump Shames a Local 'Mexican' Judge as Protestors Storm Streets." *Washington Post.* < https://www.washingtonpost.com/news/post-politics/wp/2016/05/27/in-san-diego-trump-shames-local-mexican-judge-as-protesters-storm-streets/>

the ascriptive traits of its agents, this undermines the ability of legal reforms to shore up support for the rule of law. This leads us to ask, To what extent do the ascriptive traits of legal authorities have the capacity to influence public perceptions of impropriety in the rule of law, and what are the democratic implications that result? As we argue later, the answer to this question has important implications for democracy.

We examine this question by focusing on perceptions of judicial impropriety in the United States. Judges represent some of the most visible actors in the U.S. legal system yet increasingly find their impartiality questioned, a critical fact since perceptions of impartiality help to distinguish the legitimacy of courts from that of the elected branches (Gibson 2008; Hibbing and Theiss-Morse 2002; Tyler 2003). With the judiciary becoming more diverse, ascriptive characteristics might take on greater importance, particularly if citizens are prone to doubt the impartiality of outgroup judges (Nelson 2015). We develop a theoretical framework in which ascriptive characteristics such as race and gender serve as informational shortcuts. We theorize that citizens will use these cues to develop expectations about whether a judge will behave in an improper and biased manner.

Why care about whether citizens believe judges are biased? First, when citizens perceive bias on the part of judges, this directly implicates and undercuts perceptions of procedural fairness in the legal system (Tyler 1988; 2003; Tyler and Huo 2002). Procedural fairness is a foundational resource for the effective rule of law. Indeed, "people's willingness to accept the constraints of the law and legal authorities is strongly linked" to fairness evaluations (Tyler 2003, 284). Moreoever, citizens' fairness judgments affect general supportive values, such as the institutional legitimacy of courts, decision acceptance, and compliance (Tyler 2006). Compliance by the public cannot be taken for granted, and "declining confidence in law and legal authorities may lead to declining feelings of obligation to obey … raising the possibility that compliance may be increasingly problematic" (Tyler 2003, 291). While we expore this issue further in the coming pages, we note here that our focus on

bias and impropriety is motivated directly by research showing that when people perceive courts and judges as biased, this can undermine perceptions of fairness, legal system legitimacy, and even compliance with the law.

We leverage two survey experiments – using complementary vignette and conjoint designs – to test our theory. Building on classic naming experiments (e.g., Bertrand and Mullainathan 2004), we subtly manipulate the race and gender of judges assigned to a pending case by varying only their names (e.g., Brad Sullivan versus Ariana Hernandez) in a news vignette. Importantly, we hold all other information constant, which allows us to isolate how the ascriptive attributes of judges influence citizens' perceptions of judicial impropriety. To complement our vignette experiment, we administer a conjoint experiment, enabling us to isolate the effects of a judge's race and gender when compared to many other attributes such as legal training, experience, and party affiliation.

We find that political predispositions (partisan and ideological identities) of citizens interact strongly with ascriptive judge characteristics to shape perceptions of judicial impropriety. Specifically, members of the political left are more likely to evaluate white male judges negatively, while some on the right exhibit greater skepticism toward female and Hispanic judges. The finding upends the argument that race and gender-based attacks on judges backfire (see Van Hall 2017). In fact, it suggests that such attacks may *increase* perceptions of judicial impropriety under some conditions. We also find evidence that partisan identity has a much stronger effect on support for female and Hispanic jurists than it does for white male judges.

Additionally, we demonstrate that race and gender cues have a more powerful effect for some issues. Specifically, citizens perceive more impropriety among female judges in abortion cases and among Hispanic judges in immigration cases. This finding suggests a pernicious appeal to the accusation that Hispanic judges cannot fairly adjudicate immigration cases. Importantly, we isolate a key mechanism that underpins perceptions of improper behavior: citizens' expectations about in-

group favoritism. Because our conjoint analysis allows us to independently examine multiple judge attributes, we are able to show that a judge's race and gender have some of the most powerful effects on perceived impropriety, second only to partisan affiliation.

Besides contributing a rich new source of data on public perceptions about proper behavior for judges, our findings have numerous implications. Most existing work focuses on how citizens apply demographic cues concerning lawmakers and executives. Yet because the effect of ascriptive traits is likely to vary substantially across contexts (Eagly and Karau 2002), we contribute new insights regarding an institution, the judiciary, that is becoming increasingly diverse. Our findings have direct implications for elected judges, but they also speak to critical questions surrounding public support for courts, acceptance of their decisions, and compliance with controversial rulings. In fact, our results imply that there may exist an appetite for race and gender-based attacks aimed at judges. One implication of this is that increasing diversification of the bench might polarize support for the rule of law. This danger is particularly acute when political elites attack the integrity of judges based on race or gender. But aside from publicizing these risks, our study does not offer any ready-made solutions to the problem. Although reformers may aim to make legal procedures more fair, they do not have the capacity to alter ascriptive traits.

Judicial Diversity, Bias, and Support for the Rule of Law

Recent attacks on judges raise concerns about improper behavior but specifically reference ascriptive traits as its source. Attacks like the one from then-candidate Trump have come about as diversity in the U.S. judiciary increased substantially. Today, women hold slightly more than one in four seats on the federal bench, while racial and ethnic minorities hold about 10% of slots, according to the Federal Judicial Center. Diversity on state courts varies, but it is not uncommon for one of three seats in a state to be filled by a woman (Reddick, Nelson, and Caulfield 2009). When it comes

to minorities, a few states have none on the bench, but many others are composed of at least 15% minority judges.[2]

Diversification has numerous positive aspects. For example, diverse judges bring distinct informational and experiential perspectives to the courts on which they serve, which diffuse to their colleagues (Boyd, Epstein, and Martin 2010; Glynn and Sen 2015; Kastellac 2016).[3] When judicial decisions are informed by distinct experiences and personal characteristics, this may have many benefits, including adding nuance and perspective to the law. According to Justice Sotomayor, "a different perspective can permit you to more fully understand the arguments that are before you and help you articulate your position in a way that everyone will understand" (Peltz 2016). Where it becomes more concerning, however, is if *the public believes* that judges with certain ascriptive traits are inherently "biased."

There are multiple reasons that beliefs about bias should be taken seriously. First, research on procedural justice demonstrates that believing judges and courts to be fair and unbiased is a key ingredient that shapes whether the public supports their decisions and sees them as legitimate (Tyler 2006; Tyler and Huo 2002). This can also extend into compliance. Courts and judges lack the power to enforce decisions and, while compliance is commonplace, it is not guaranteed (Tyler 2003). Of course, merely accusing judges of bias or impropriety may not be enough to undermine the public's belief in their fairness, so a key question of this study is whether Americans are predisposed to believe that some judges, namely women and minorities, are uniquely infected by bias.

---

[2] Diversity data from the American Bar Association's National Database on Judicial Diversity in State Courts. <http://apps.americanbar.org/abanet/jd/display/national.cfm>

[3] To be clear, there is also substantial within-group variation in the behavior of women and minority judges.

Second, if some citizens are predisposed to seeing certain judges as biased, this can make it more difficult for women and minorities to be elected to the bench. Currently, 21 states elect their high court judges and 19 others subject them to retention elections, and research suggests that gender affinity has a key influence on vote choice in these elections, particularly in nonpartisan contexts (Badas and Stauffer 2019). At the federal level, judges do not directly face voters, but this does not imply that popular support is unimportant. This is because presidents take into account potential public support for judicial nominees and public support can also influence the likelihood of Senate confirmation (Kaslovsky, Rogowski, and Stone 2019; Kastellac, Lax, and Phillips 2010). Returning to the Sotomayor example, some Republican senators attributed their votes to reject to her off-bench comments regarding her ethnicity.[4] In short, if Americans are predisposed to doubt the fairness of judges based on ascriptive characteristics, this has implications for the makeup of the bench.

Third, whether or not some Americans believe female and minority judges to be "biased," if these judges have reason to *believe* that they will be uniquely scrutinized it can have implications for their decisions.[5] Female and minority judges may be forced to go to greater lengths to write high quality opinions, exhibit judicial independence (Choi et al. 2011), and avoid reversals (Sen 2015) than white male judges. Female judges exhibit more independent behavior than male judges (Choi et al. 2011). The idea that bias against women and minority judges leads them to adjust their behavior echoes a finding from the congressional literature, which shows that female lawmakers outperform

---

[4] See Means, Bill. 2009. "Two Key GOP Senators to Oppose Sotomayor." CNN.

<https://www.cnn.com/2009/POLITICS/07/24/sotomayor.vote/index.html>

[5] Research suggests that judicial performance ratings "yield biased results based on gender and race" (Gill, Lazos, and Waters 2011, 733; Gill 2014).

their male counterparts in terms of legislative productivity (Anzia and Berry 2011). In short, if female and minority judges believe that some citizens will see them as biased or unqualified, this has implications for the opinions they write, their interactions with colleagues, and even the conditions under which they dissent (Choi et al. 2011).

Fourth, since we are focusing on public opinion about judges, it is necessary to acknowledge the limits of Americans' awareness when it comes to courts. Most citizens are not paying attention to most decisions reached by most judges (Gibson and Caldeira 2009; Jamieson and Hennessy 2006). However, Trump's attacks on federal judge Gonzalo Curiel indicate that on the occasions courts garner substantial public attention, some of the most readily available information citizens have at their disposal concerns the race and ethnicity, gender, and (potentially) partisan identification of judges. These four reasons make it critical to explore the effect of ascriptive traits on Americans' perceptions of judicial bias.

Evaluating Judges on the Basis of Ascriptive Traits

Existing research conceives of race and gender cues as an informational shortcut, or heuristic, that citizens can use to reach judgments about political affairs. To do so, citizens extract information from the cue and make fast and frugal judgments. Race and gender cues serve at least two functions, calling up pre-existing group stereotypes in the minds of citizens and signaling an actor's ideological predispositions. This need not be conscious: cues may activate schema through an automatic, involuntary process (Devine 1989).

Although many citizens do not display negative attitudes towards women in politics (Dolan 2014; Sanbonmatsu 2002), a nontrivial portion of the public believes women are poorly suited to politics (Burden, Ono, and Yamada 2017; Dolan 2010; Lawless 2004). Role congruity theory suggests that anti-women attitudes are more likely to manifest when women seek or hold political offices whose requisite skills do not align with the perceived "strengths" of women (Eagly and

7

Karau 2002; Huddy and Terkildsen 1993; Lawrence and Rose 2014; Rose 2013). Some citizens, for example, oppose women wielding executive power because they believe that women lack the requisite decisiveness.

Similar patterns are evident for minority candidates. From an ingroup favoritism perspective, voters tend to prefer candidates who share their racial background, which can advantage white candidates in contests where their co-ethnics constitute a majority of the electorate (Barreto, Villarreal, and Woods 2005; Huddy 2001). More dramatically, prejudice also plays a role in shaping political preferences, with negative consequences for minority candidates (Sides, Tesler, and Vavreck 2018; Tesler 2012). In spite of these broad patterns, an actor's racial background can have nuanced effects. Many citizens hold strong norms of fairness, meaning that explicitly racialized appeals, and even some implicit ones, may not harm minority candidates (Tokeshi and Mendelberg 2015). In addition, voters use ascriptive traits to infer the policies or groups a candidate favors, which can supply electoral advantages to female and minority Republicans (Koch 2000; Koch 2002; Meyer and Woodard 2017). This is because voters assume that women (McDermott 1997; Hayes 2011) and minorities (McDermott 1998; Valentino and Hutchings 2004) prefer more liberal policies. Citizens may therefore perceive minority and female Republicans as more ideologically moderate since diversity (or non-white male identity) signals they are not as conservative as a typical member of their party. Additionally, citizens make inferences about politicians' specific issue positions on the basis of their demographic traits. Female politicians are commonly seen as promoters of women's issues (e.g., strong positions against sex-based discrimination) and other matters that are typically associated with the political left.

In short, citizens perceive demographic cues as providing information. The value of this information varies dramatically across contexts, voters, and candidates. Yet existing work focuses on how citizens apply demographic cues concerning lawmakers and executives. Because the

application of demographic stereotyping is likely to vary substantially based upon the office involved (Eagly and Karau 2002), there are important unanswered questions when it comes to how they shape judgments about judges and the legal system more generally.

Ascriptive Traits, Perceptions of Impropriety, and Judgments about the Legal System

Our account explains the conditions under which male versus female and white versus minority judges are susceptible to charges of impropriety and bias. Such charges are politically important because they call into question the authority of a diverse judiciary, with implications for how judges behave. While our approach builds on insights from research on heuristic processing, our theory goes beyond research on voting behavior by speaking to how the effects of race and gender differ across distinct legal issues.

We begin from the premise that citizens pay limited attention to judicial affairs. Attention to Supreme Court is modest and, naturally, information about the judiciary becomes even more limited when it comes to local, state, and lower federal courts (Vining and Wilhelm 2011). Yet citizens do encounter news about lower court decisions from time to time in brief pieces on local news and internet. As shown in a wealth of research, when faced with only snippets of information, citizens tend to rely on cognitive heuristics to arrive at judgments (e.g., Lupia 1994; Lupia 2016). Thus, gender and race/ethnicity are likely to serve as cues for making inferences about an actor's preferences. We expect this same tendency to extend to evaluations of judges (Nelson 2015).

How might citizens use demographic cues? We posit a link between the perception that a judge favors a particular outcome and the broader judgment made by citizens about whether that judge is biased. To be clear, we are not arguing that citizens use race and gender cues solely for the purposes of evaluating a judge's ideological preferences. Citizens hold realistic conceptions of judicial behavior, and believing that judges have ideological preferences is not necessarily harmful to legitimacy (Gibson and Caldeira 2011; Gibson and Nelson 2017). However, we argue that

9

stereotyping can prove damaging if citizens believe that judges behave in a *prejudicial or biased* fashion, favoring certain litigants such as in-group members.

Research demonstrates that citizens perceive a close connection between certain issues and demographic groups (Fridkin and Kenney 2009; Lawless 2004). For example, cases involving pay discrimination or abortion may be perceived as having gender-based implications, while those involving immigration may be associated with ethnicity. We anticipate that citizens recognize these distinctions when evaluating judicial fairness. We focus on two issues in particular: abortion rights cases, which citizens often perceive as a "women's issue," and immigration, which ties closely to views about Hispanic Americans. At the same time, we acknowledge the fact that stereotyping is not unique to female and minority judges, for it is also possible that citizens will perceive white male judges to favor parties that share their traits.

When citizens perceive the potential for judges to favor their in-group, this presents a more severe violation of fairness than simply deciding in light of ideological preferences. The Judicial Code specifies a need for disqualification in cases where a judge's impartiality could be reasonably questioned. Yet because citizens process information for instrumental purposes, they are likely to express the most concern about judicial bias when a judge favors groups about which they feel negatively. We expect that for members of the political right women and minorities that display in-group favoritism are evaluated in particularly negative terms.

> *Gender hypothesis*: All else equal, members of the political right (left) should be more (less) likely to view female judges as prejudicial.
>
> *Women's issues corollary*: All else equal, citizens should be more likely to apply gender (versus race) cues in cases involving quintessential "women's issues." Specifically, women's issue cases should *increase* the effects of gender, relative to race, on the gap in perceived judicial prejudice between members of the political left and right.

10

Ethnicity hypothesis: All else equal, members of the political right (left) should be more (less) likely to view Hispanic judges as prejudicial. [6]

Hispanic issues corollary: All else equal, citizens should be more likely to apply race (versus gender) cues in cases involving quintessential "Hispanic issues." Specifically, Hispanic issue cases should *increase* the effects of race, relative to gender, on the gap in perceived judicial prejudice between members of the political left and right.

To summarize our theoretical insights from above, although Americans have limited knowledge about the judiciary, determining its overall fairness is a key consideration. Citizens will draw on the most readily available information when evaluating the fairness of judges, including cues about their race and gender. Doing so has systematic implcations for judgments about whether a judge can behave in unbiased fashion. This is because some citizens will tend to believe that judges display "bias" by favoring parties that share their race or gender. Moreover, these patterns should strengthen when it comes to issues with the potential for strong in-group favoritism by judges. Our framework implies that the increasing diversification of the bench may increase polarization over the performance of judges.

Study 1: Vignette Experiment

Our first study consists of a survey experiment using a sample (N=3,117) of the American voting age public recruited by Survey Sampling International in summer 2018. SSI balanced respondents for this study by age, gender, ethnicity, and census region. Potential participants were contacted with notifications using email and, once within SSI's system, participants were matched with an available survey with multiple points of randomization. Respondents were replaced for clear evidence of extreme "satisficing" behavior, including straight-lining and speeding as well as failing two attention check questions. The appendix describes sample attributes in more detail on page 1.

---

[6] For ethnicity, we focus on Hispanics because they represent one of the most visible minorities in the U.S. and Hispanic judges have been attacked for displaying favoritism in certain cases.

Our embedded experiment presents realistic information about pending cases assigned to judges in the U.S. court system. To maximize external validity, we created vignettes based on actual news coverage across two issue areas, abortion rights and immigration. This is important because we expect an interaction between case type and ascriptive traits, with the impartiality of female judges under greater scrutiny in the former and Hispanic judges in the latter.

While vignette studies of judicial decisions are commonplace, we introduce a key innovation. Within the vignette, we subtly manipulate a judge's race and gender through random name and pronoun assignment while keeping all other information constant. We used six potential name categories, randomly assigned and detailed in Table 1. We drew on existing research about name perceptions to generate names that Americans are most likely to perceive as Caucasian or Hispanic (Bertrand and Mullainathan 2004; Gaddis 2017). In addition, we also inserted names stereotyped as African American in order to conduct a "placebo" test: since our case vignettes focus on abortion and immigration, we expect strong effects for gender and ethnicity but not for race. Full experimental stimuli are presented in the appendix.

Table 1: Priming Judicial Traits

Judge name	Trait indicator	Manipulation check (% perceiving "correct" race)
Brad Sullivan	White man	88.1
Anne Sullivan	White woman	84.5
Diego Hernandez	Hispanic man	84.2
Ariana Hernandez	Hispanic woman	80.7
Darnell Washington	Black male	58.3
Tamika Washington	Black female	72.1

The careful reader may wonder whether the name priming was too subtle to influence awareness of the judge's race and gender. This was not the case. A manipulation check shows that respondents overwhelmingly assigned the judges the stereotypical racial classification associated with their name. 78% "correctly" matched the judge to his/her race. Moreover, assignment to the

Hispanic judge condition significantly increased the likelihood of believing the judge to be Hispanic (p<0.001).[7] In the main models below, we analyze the results for the full sample, which gives the effect of the intention to treat (ITT). However, we also estimate the treatment on the treated by analyzing only those who correctly answered the manipulation check. We find very similar results.

We used a multi-item scale to capture perceptions of judicial impropriety, with five-point Likert response options for each. First, we asked respondents to rate whether the judge would "display improper bias when ruling" on the case. This bias question gets to the heart of recent accusations of partiality and also has implications for procedural fairness, so it is a key cog in our theory. Additionally, we leverage the idea that perceived conflicts of interest increase the likelihood of recusal, asking whether the judge "should be required to recuse [himself/herself] from the case" (see Gibson and Caldeira 2012). Finally, we asked whether the judges' "values and political views will influence" their rulings. We recognize that this final item is not negative on its face. We include it, however, based on the literature that demonstrates that citizens extend less legitimacy to courts and judges that they perceive to be influenced by non-legal considerations (Christenson and Glick 2015). In addition, we note a strong correlation across our items when we examine their psychometric properties, suggesting that beliefs about bias, recusal, and political influence are tapping into the same underlying construct of *improper* behavior for judges. We include analysis in the appendix that focuses on the bias item alone, finding very similar results to the ones we present here.

Respondents varied widely in the impropriety they perceived. On the bias item, 11.3% of respondents strongly disagreed with the possibility of bias, but 20.7% agreed and another 8.7%

---

[7] More than 7 in 10 respondents matched white male, white female, black female, Hispanic male, and Hispanic female judges with the "correct" race. The black male judge manipulation was less successful (58.3% "correct").

agreed strongly. On the recusal item, 37.6% did not advocate recusal (disagree/strongly disagree) while 24.6% did, and another 37.8% were neutral. Finally, on the values item, the modal respondent agreed that judges were influenced by personal views and political values (35.3%), while 12.2% strongly agreed and 5.9% strongly disagreed. Notably, responses to all three items scaled together when subjected to factor analysis, with loadings at 0.78, 0.76, and 0.64 respectively. The eigenvalue of the single significant underlying factor, which we assume to capture the latent dimension of improper bias, is 1.60. We focus on this factor score in subsequent analysis of our vignette experiment, although note that the results are substantively similar if we analyze the items independently.

We analyze how the randomly assigned conditions (issue area, gender prime, and race/ethnicity prime) interact with political predispositions, which we measure by combining responses to the standard 7-point partisan identification and ideology questions (r=0.63). This is advisable due to the increasing correlation of partisan and ideological identities in the U.S. (Mason 2015; Mason and Wronski 2018). Therefore, our combined left-right identity measure ranges from 0 (very liberal, strong Democrat) to 12 (very conservative, strong Republican). However, we note that the results are very similar if we isolate the partisan or ideological measure in our specification. In all models, we include pre-treatment covariates for respondents' race, ethnicity, education, and political knowledge, measured using three items.

Results of Study 1

We begin by testing our expectation that citizens will use race and gender cues to infer whether judges are prejudiced. We explore this using a question to measure whether respondents expect the judge to rule in a pro-abortion or pro-immigrant fashion. We display the logit model estimates in Table 2 and find support for our expectations for gender and ethnicity.

Table 2. Likelihood the Judge is Predisposed to Favor a Liberal Outcome

Randomly assigned	Coefficient estimate
Female judge condition	0.27**
	(0.08)
Hispanic judge condition	0.47**
	(0.10)
African American judge condition	0.14
	(0.10)
Pseudo-R^2	0.09
N	3153

Note: Model using a logistic specification, with standard errors in parentheses. *p<0.05, **p<0.01. Comparison condition is a white male judge. Model also controls for assignment to case conditions as well as a respondent's gender, race, ethnicity, education, PID, ideology, and knowledge.

First, random assignment to a female judge condition increases the belief that the judge will favor a liberal outcome. In substantive terms this equates to about a 6-percentage point increase, from about a 38% to 44% likelihood of the liberal group winning, all else equal. While not a massive effect, remember that it comes about from a very subtle treatment – random assignment to distinct names and pronouns. Moving on to race and ethnicity, the subtle name manipulation also has the anticipated effects. All else equal, Hispanic judges are perceived as much more likely to favor liberal outcomes (expected by 46.5% of respondents) than their white counterparts (expected by 36.3% of respondents).

Another notable aspect from Table 2 is consistent with our theory: we find null results for the African American judge. If citizens were simply stereotyping all minority judges as liberal, this would not be the case. Rather, we have hypothesized that respondents will perceive only female and Hispanic judges as more liberal since we have primed abortion and immigration cases in our study. This is just what the results in Table 2 suggest.

Having established that citizens stereotype female and Hispanic judges as more likely to favor certain groups, and white male judges as less likely to do so, we next examine our hypothesis that judge traits will interact with political predispositions to influence whether respondents believe judges will behave in an improper fashion. This is important because while voters may expect

political candidates to favor certain policies, they may *penalize* judges for privileging certain outcomes or parties ahead of time. We model perceptions of improper bias in Table 3.

We are particularly interested in the interaction between the treatment and a respondent's political predispositions, and we find consistent support for our expectations. In the pooled results, the estimated coefficients for the interaction terms show that liberal Democrats and conservative Republicans diverge in their assessments of impropriety based on judge gender and ethnicity. The interaction is significant ($p<0.05$) and positively signed for both female and Hispanic judges, indicating that as we move from liberal Democrats to conservative Republicans, the perception of judicial impropriety increases when a judge is a woman or Hispanic (in comparison to their white/male counterparts). Not only do respondents stereotype judges on the basis of ascriptive traits, but they specifically view certain judges as more biased.
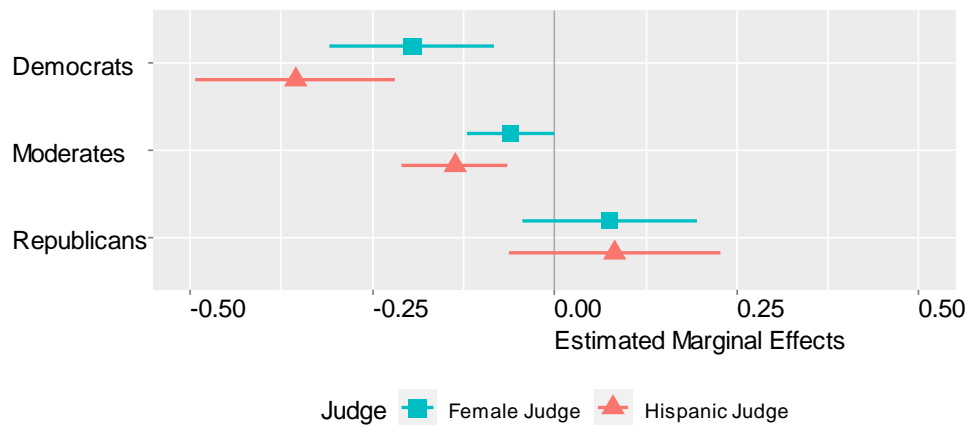
Table 3. Perceptions of Judicial Impropriety			
	Pooled Results	Separated Results	
		Abortion case	Immigration case
Female judge condition	-0.20**	-0.14	-0.25**
	(0.06)	(0.08)	(0.08)
Female judge condition *Left-right predispositions	0.02**	0.02	0.03*
	(0.01)	(0.01)	(0.01)
Hispanic judge condition	-0.36**	-0.23*	-0.48**
	(0.07)	(0.10)	(0.10)
Hispanic judge condition *Left-right predispositions	0.04**	0.01	0.06**
	(0.01)	(0.01)	(0.01)
Left-right predispositions	-0.01	-0.005	-0.02
	(0.01)	(0.01)	(0.01)
R^2	0.06	0.04	0.06
N	3117	1568	1549

Note: Dependent variable is the factor score that combines the bias, recusal, and values items. Models use an OLS specification, with standard errors in parentheses. *$p<0.05$, **$p<0.01$. Comparison condition is a white male judge. Models also control for assignment to case condition, to the African American judge condition and its interaction with predispositions (insignificant in all cases), and a respondent's gender, race, ethnicity, education, and knowledge.

We probe these results further from a substantive perspective in Figure 1. Here, we display variation in the predicted level of judicial impropriety as a result of ethnicity, gender, and political predispositions using the pooled model. The gender effects are shown by the squares in Figure 1. We find strong evidence of polarization when it comes to how citizens evaluate the bias of male versus female judges. When randomly assigned to the female judge condition, liberal Democrats believe the judge to be very unlikely to exhibit impropriety – about 20 percentage points below the male judge level. This means that perceived impropriety drops from about the 58th percentile to about the 36th percentile on the scale. The pattern also holds for moderates, but the effects are less pronounced from a substantive perspective. On the other hand, Republicans are no more likely to perceive impropriety in the behavior of female judges, across all cases, as the confidence intervals on the point estimate overlap 0. We note, of course, that the inverse is also true: liberal Democrats see male judges as *more* biased, by about 20 percentage points.

Figure 1. Perceptions of Impropriety Based on Innate Traits

Note: Figure 1 shows the estimated marginal effects of changing from a white male to a female judge (squares) or a Hispanic judge (triangles) on perceptions of impropriety, using model estimates displayed in Table 3. Substantively, this demonstrates partisan polarization over judicial performance on the basis of ascriptive traits.

We turn next to the effects of judge ethnicity, shown by the triangles in Figure 1. Once again, Democrats perceive Hispanic judges as unlikely to display impropriety. In fact, the effect of

ethnicity alone is enough to change perceptions by over 35 percentage points. Perceived impropriety drops from about the 59th percentile to about the 32nd percentile on the scale. On the other hand, Republicans are more skeptical as to whether Hispanic judges can rule fairly, although the effect falls short of significance. If a judge is Hispanic, our model estimates that this increases the likelihood, by about 11 percentage points, that Republicans see the judge as exhibiting improper behavior. This also implies that Republicans see white judges as marginally more fair than Hispanics, while Democrats rate them as less fair. Moreover, as we show in the appendix, these patterns occur across multiple separated DVs, including the bias item alone and the recusal item alone. In short, we find clear evidence that Republicans and Democrats polarize over whether a judge can behave in an unbiased fashion on the basis of her race and gender.

To examine our issue-specific hypotheses, we turn to the separated models in Table 3 (labeled "Abortion case" and "Immigration case"). We expect that citizens will polarize more strongly over the performance of women, as opposed to Hispanic, judges in the abortion case, but that this pattern will reverse for the immigration case. We find modest evidence in line with our expectations. The coefficient on the Hispanic*predispositions interaction is both significant and substantively larger when an immigration issue is at stake. On the other hand, the coefficients for gender are similar across case type, indicating that citizens do not necessarily believe female judges are more biased when abortion issues are at stake. Overall, we have reason to be cautious when interpreting our issue-specific results, but they suggest that the intersection of race and gender may also have a conditioning role. For example, conservative Republicans may perceive Hispanic women as the most biased decision-makers of all. This possibility unlocks further implications when it comes to identity politics and in-group bias that we consider further in the appendix. Our analyses there also help us to better understand the partisan asymmetries that appear in Figure 1. Next, we

discuss a distinct experimental approach to further isolate the component-specific effects of judges' race and gender.

Study 2: Conjoint Experiment

To build on our main results, we conducted a second study utilizing a conjoint experimental design that isolates the effects of particular characteristics on attitudes (Hainmueller, Hopkins, and Yamamoto 2014; Sen 2017). The design operates by presenting respondents with randomized, distinct judges' profiles in which core attributes vary.

In 2017, we recruited U.S. adult respondents from Amazon's Mechanical Turk marketplace, following best practices to ensure high quality data (Berinsky, Huber, and Lenz 2012; Christenson and Glick 2013). Specifically, we ensured that respondents had participated in over 1,000 tasks on mTurk with an over 95% approval rating for their work, and we excluded 147 respondents who failed to complete and 91 respondents who shared identical IP addresses. This leaves us with 2,950 respondents in our sample. In keeping with patterns that are common to mTurk, our sample skews a bit more white and young than the general population. We subsequently present results from our unweighted mTurk sample, though the results are similar when the sample is weighted to match the demographic breakdown in the 2010 census.

We randomly assigned respondents to receive a description of a pending case, adapted from actual news coverage and concerning similar issues to study 1. The first case featured a lawsuit brought by pro-life activists to prevent the opening of an abortion clinic, while the second case concerned charges brought against an undocumented immigrant following an assault. The appendix presents full stimuli. After reading the vignette, respondents were asked to compare between ten pairs of judges who may rule on the dispute. This set up, the essence of the conjoint design, allows us to randomly vary multiple characteristics of the judges' profiles to isolate the effects of any one. The profiles we presented randomly varied *gender, race/ethnicity, age, party affiliation, legal experience, law*

19

school ranking, *marital status*, and *parental status*. The full set of randomized traits is presented in Table

4. Table 5 shows one hypothetical pair of judge profiles that one of our respondents has seen.

Table 4. Possible Judge Profiles Presented to Respondents

Attribute	Potential Traits
Gender	Male, Female
Race/ethnicity	White, Hispanic, Black, Asian American
Party affiliation	Democrat, Republican
Age	44, 52, 60, 68, 76
Legal experience	No experience, 5 years, 10 years, 15 years, 20 years
Law school ranking	Top 10 (Tier 1), 50-100 (Tier 3), 151-200 (Tier 4)
Marital status	Single, Married
Parental status	No children, 1 child, 2 children

Table 5. Example Pair of Judge Profiles Presented to Respondents

	Judge A	Judge B
Gender	Female	Male
Race / ethnicity	Black	White
Party affiliation	Republican Party	Republican Party
Age	76 years	68 years
Experience in legal profession	20 years	15 years
Law school ranking	Top 10 (Tier 1)	50-100 (Tier 2)
Marital status	Married	Married
Parental status	1 child	No children

There are a few things to emphasize about this design. First, because we vary multiple

attributes simultaneously, we are able to isolate the effect of any one attribute independent of the

others. This is known as the average marginal component-specific effect (AMCE). So, for example,

different from our vignette experiment, we can isolate the influence of a judge's gender *independent* of

his or her partisan affiliation even though gender may sometimes be taken as a signal about party

preference. We can also explore how attributes interact with respondent traits, such as gender and

partisanship. This is known as the average component interaction effect (ACIE), and these results

are also independent of all other attributes. Additionally, the design retains a high measure of

external validity. We presented respondents with descriptions of cases adapted from actual civil and

criminal law disputes. Also, we have designed studies 1 and 2 to complement one another. In the former, we subtly manipulate judges' race and gender cues using name primes alone. In the latter, we explicitly manipulate race and gender among a variety of other attributes. Therefore, while study 1 presents information in a fashion akin to what citizens may encounter in actual news coverage, study 2 provides significant control over many judicial attributes, allowing us to isolate the effects of any one component in light of multiple others. As we show, these distinct methods produce very similar results.

To evaluate improper bias, we followed up each judge-pair by asking respondents to evaluate the judges along two dimensions. First, respondents rated which judge was "more likely to display improper bias when ruling on the case." Second, respondents evaluated which judge they believed was "more likely to have their values and political views influence how they decide."[8] Since both

---

[8] By using the forced choice design, we encouraged respondents more carefully evaluate the judge profiles. If respondents believe that judges decide primarily on the basis of law, then they may randomly select one of the judges, considering that neither of the judges is likely to have their views shaded by non-legal factors. Since judges' attributes are randomized, such behavior does not produce biased outcomes. Indeed, Hainmueller, Hangartner, and Yamamoto (2015) show that the results in a paired conjoint with forced choice quite accurately match the behavior found in the real world. Moreover, if everyone has no bias against certain types of judges, we should find null results (i.e., random choosing would bias against our findings). There might be a concern that the repetitive forced-choice taks of conjoint experiment in the survey would lead survey satisficing among respondents. While we did not allow respondents to skip items, including those of our conjoint experiment, only 147 out of 3,188 (4.6%) failed to complete our survey. 71 out of 147 (48%) terminated the survey before reaching conjoint experiment items, and 76 out of 147 (52%)

results display substantively the same patterns of outcomes, we present only the results based on the responses to the bias question for the sake of space, and show the results drawn from the second question in the appendix.

Results of Study 2

Our dependent variable is dichotomous so we use a logistic specification and cluster standard errors by respondent, which accounts for the fact that each respondent saw multiple judge pairs (Hainmueller, Hopkins, and Yamamoto 2014). We evaluate our hypotheses by exploring the ACIEs in Figure 2. To reiterate, the coefficient estimates represent the effect of the *interactions between a respondent's party identification and the judge's trait* named in each row. Figure 2 demonstrates that a respondent's partisanship interacts with judicial attributes in multiple ways. The results show that Democrats and Republicans assess judges very differently. Democrats rate female judges as much less likely to behave improperly than their male counterparts when ruling. Specifically, compared to otherwise identical male judges, women are rated by Democrats as approximately 6% less biased.[9] Yet these patterns reverse for Republicans, who see female judges as about 2% more likely to fall prey to bias.[10] This partisan pattern provides further evidence for our gender-partisanship hypothesis.

For minority judges, we are interested in the interaction between a respondent's partisan predilections and a judge's ethnicity in Figure 2. Once again, we find clear evidence of partisan

---

terminated in the middle of conjoint experiment. Thus, respondents do not seem to quit answering our conjoint items at a substantially higher rate than others in the survey.

[9] This effect is as large as that of a judge with 5-10 years of legal experience.

[10] The magnitude of this effect is about the same as if the judge graduated from a low ranked law school.

polarization as expected. Republicans are 10% more likely to rate Hispanic judges as biased when

compared to white judges. This is notable because it represents one of the largest effects uncovered

in the experiment outside of partisanship. Where Republicans are more likely to view racial

minorities as biased judges, Democrats are less likely to, which also fits with the ethnicity patterns

from the vignette experiment. Alternatively, the pattern reverses for white and male judges, whom

Republicans are *less* likely to perceive as biased. Democrats perceive them as more biased.

Figure 2. Pooled Conjoint Results Predicting Judge Bias
Average Component Interaction Effects (judge is biased)
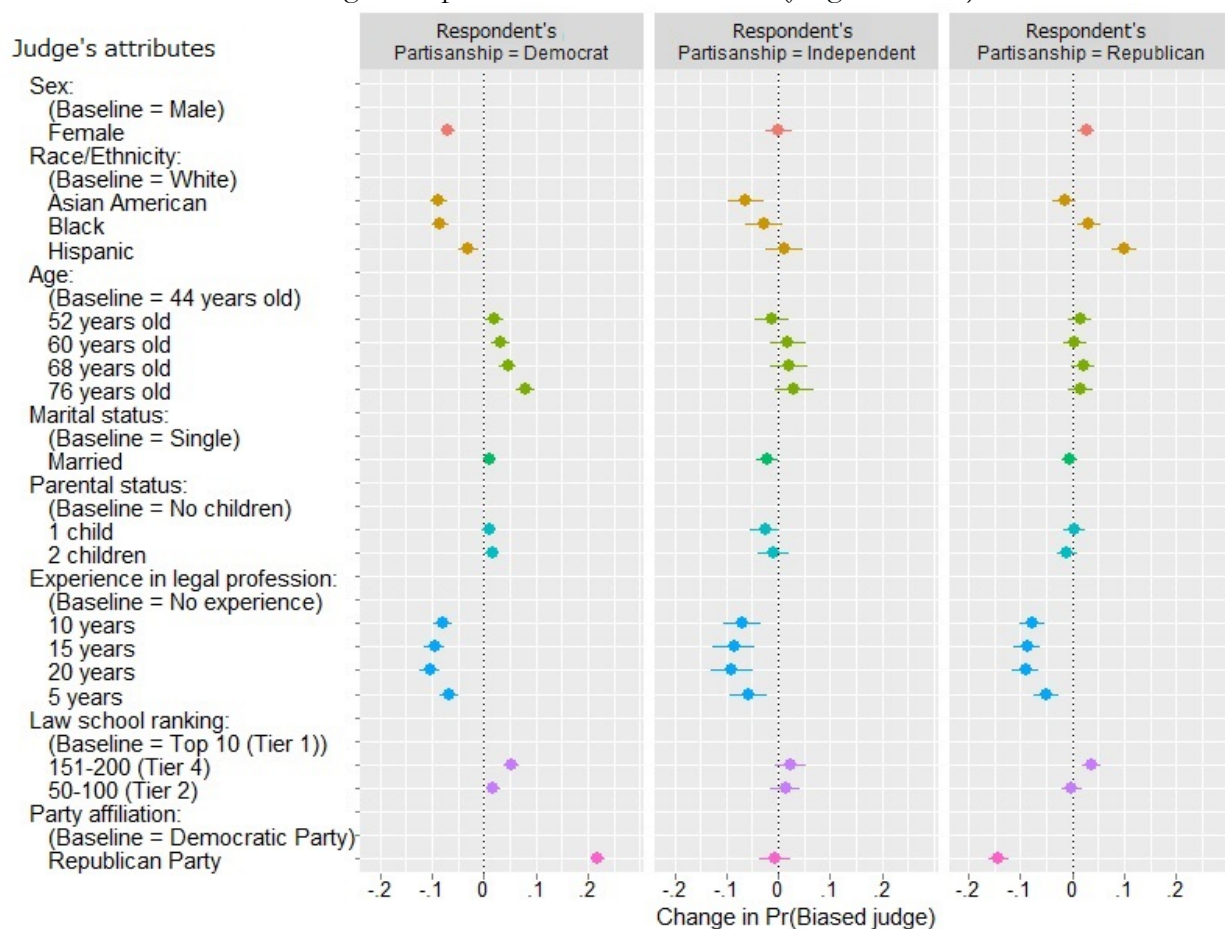
Note: Manipulated judge attributes on y-axis, respondent partisan characteristics across three panels. Figure shows the estimated effect of each judge trait on perceptions that the judge is biased, conditional on a respondent's partisanship. Substantively, this demonstrates partisan polarization based on gender and race/ethnicity, with Democrats perceiving female and Hispanic judges as less biased and Republicans perceiving them as more biased. These effects are estimated while controlling for a judge's partisanship, meaning that they are not simply a function of partisan policy disagreement.

Figure 3. Separated Conjoint Results for Abortion Case
Average Component Interactive Effects (judge is biased)
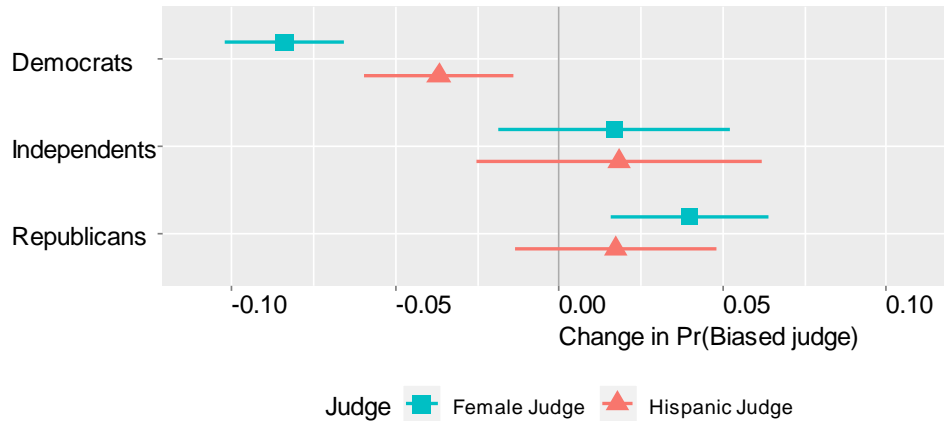
Judge   ■ Female Judge   ▲ Hispanic Judge

Note: Figure shows the estimated effect of partisanship on perceptions that the judge is biased, conditional on the judge's gender and ethnicity. Substantively, this demonstrates partisan polarization based on gender and ethnicity, with Democrats perceiving female and Hispanic judges as less biased and Republicans perceiving them as more biased. Note that these are partial ACIEs from a model that controls for all other traits, including partisan policy disagreement. See the appendix for the full ACIEs for this figure.

We find support that is consistent with our issue specific expectations, suggesting that citizens believe judges are somewhat more likely to display bias on certain types of cases. These results suggest that the mechanism of perceived in-group favoritism is powerful. With Figure 3, we explore perceptions of judicial bias in only abortion cases. This figure demonstrates a strong influence of gender on perceived bias in abortion cases. Specifically, Republicans are approximately five percentage points more likely to rate female (as opposed to male) judges as biased when they have been assigned to resolve an abortion controversy, an effect that we suggest has been brought about by the fact that Republicans expect women to unfairly favor a "liberal" position. But this is not the only evidence for this proposition, for we also observe predictable effects in the abortion case among Democrats and even Independents. Unlike Republicans, Democrats perceive female judges as particularly *unbiased* in abortion cases – about eight percentage points less likely to display bias. Put differently, the results among Democrats show that they perceive *male* judges as more biased in abortion cases. Finally, there is no evidence that self-identified Independents perceive greater bias among male *or* female judges in abortion cases.

24

But while respondents believe that jurists may pre-judge abortion cases based on their gender, other judicial attributes have a much less clear effect. To see this, we can turn to the Hispanic ethnicity results from Figure 3. Notice that there is no evidence that Republicans or Independents perceive more bias on the part of Hispanic judges in abortion cases. On the other hand, Democrats see a bit less bias among Hispanic judges in abortion cases, although this effect is more muted than it is for female judges. Specifically, Democrats rate Hispanic judges as about four percentage points less biased in abortion cases compared with white judges, but rate women as about eight percentage points less biased, all else equal. What appears to be happening here is a process by which respondents match demographic information with case specific information, determining when it is most relevant and, thus, when to apply it in order to assess judicial bias.

We find further support for this proposition when we turn to the immigration case results in Figure 4. If citizens believe that ingroup favoritism shapes the behavior of judges in only certain cases, we would expect the most pronounced results with respect to Hispanic ethnicity for immigration cases. This is exactly what we find. Notice in Figure 4 that Republicans rate Hispanic judges as about 17 percentage points more biased than white judges in immigration cases, a very strong effect.

Finally, we see further support for our expectations of a link between judge attributes and issue when we turn to the effects of gender in the immigration case. Democrats (as opposed to Republicans) still tend to rate women as a bit less biased. Yet when we compare the results for judge gender to ethnicity in the immigration case, the latter has predictably stronger effects. Figure 4 illustrates this by showing that Democrats and Republicans differ by about six percentage points in their evaluations on the basis of gender, but by about 20 percentage points based on ethnicity.

Figure 4. Separated Conjoint Results for Immigration Case

Average Component Interactive Effects (judge is biased)
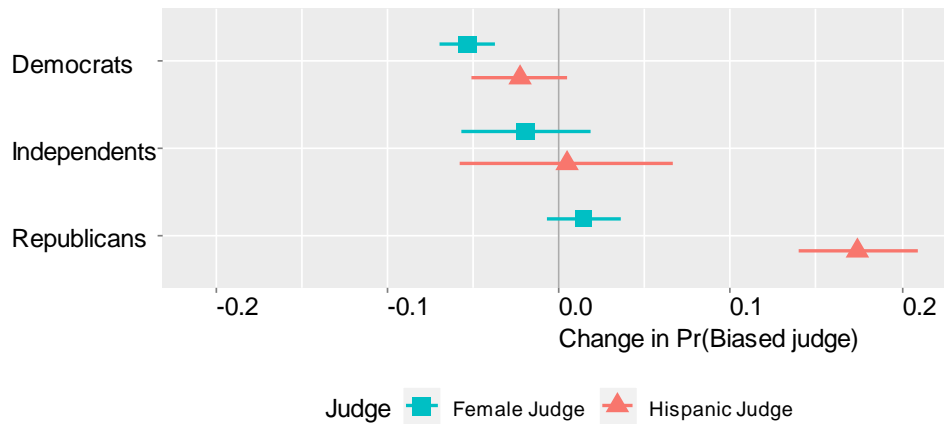
Note: Figure shows the estimated effect of partisanship on perceptions that the judge is biased, conditional on the judge's gender and ethnicity. Substantively, this demonstrates partisan polarization based on gender and ethnicity, with Democrats perceiving female and Hispanic judges as less biased and Republicans perceiving them as more biased. Note that these are partial ACIEs from a model that controls for all other traits, including partisan policy disagreement. See the appendix for the full ACIEs for this figure.

One way to understand the substance of these effects involves considering their implications when it comes to competitive elections. Many judges in the U.S. are popularly elected and must be attuned to maintaining support from a majority of constituents. Yet our results show that the gender and ethnicity of a judge shape whether a substantial number of citizens perceive them as unbiased. In close elections, a swing of even a few percentage points matters a great deal, so the implications of our results for elected judges are substantial.

Overall, our findings show that in spite of the progress they have achieved, female and Hispanic judges face remaining hurdles. Equally notable is the fact that ascriptive traits contribute further to polarization in support for U.S. judges. This pattern is concerning because it goes beyond mere ideology, demonstrating that citizens believe that judges are biased in favor of litigants who share their ethnicity or gender. As women and minorities make up a larger share of the bench, our results imply that partisans diverge in whether judges can rule without bias. As another consideration, readers may be interested in how the effects we uncover differ depending upon

respondents' backgrounds (race, gender, etc.). In the appendix, we present separated ACIEs and discuss the differences in detail, finding some interesting patterns.

Discussion

Despite significant progress, women and minorities continue to face barriers in politics and the law. Research indicates that there is skepticism among a subset of the population when women and minorities run for elective office. Yet we demonstrate here that citizens wield an influence that can negatively impact marginalized groups even after they have taken office. Specifically, we find that some citizens continue to raise questions about the capabilities of female and minority judges as they rule. In keeping with concerns about the influence of political polarization today, our results suggest a dramatic division in opinion between Democrats and Republicans depending upon the ascriptive traits of judges.

Our work adds nuance to research on support for judges and judicial legitimacy. Simply put, some citizens stereotype women and minority judges as more likely to pre-judge a case in favor of a member of their in-group. We also demonstrate the conditional nature of this effect, showing that it occurs most readily in cases in which citizens perceive a tie between a judge's background and the issue on which she is to rule. Because the perceptions that judges are fair and unbiased is a key ingredient in shaping judicial legitimacy (Tyler 2003), the implications of our results may not be isolated to a specific judge. In addition, our results have direct implications when it comes to elected judges in the U.S., suggesting additional hurdles faced by female and minority jurists.

The use of demographic cues has reverberations beyond the legal system. In and of itself, it is not too concerning that women and minority judges are seen as a bit more liberal. But it is alarming when citizens say these groups are *more likely to be exhibit improper bias as a result*. Charges of impropriety cut at the heart of the judiciary's support. Judges depend on the perception that they are professional, disinterested decision-makers to ensure support for their institution and

27

compliance with their rulings. Our findings, therefore, speak to fundamental questions about the rule of law in the U.S. and also square with the increasing attention to how identity politics infects multiple aspects of the political system (Kinder and Kalmoe 2017; Jardina 2019).

To be sure, the story is nuanced. Even something as simple as support for rulings depends upon a variety of case-specific and contextual factors. But it is important to recognize the central role of race and gender in the judicial context. Existing studies that emphasize the contextual nature of support often focus on the more information-rich environment surrounding the U.S. Supreme Court (e.g., Baird and Gangl 2006). But when it comes to most courts, citizens' knowledge may be limited to information about a judge's gender, race, or party affiliation. Trump's attacks on U.S. District Judge Gonzalo Curiel are instructive. The vast majority of citizens might only have known the judge's gender, ethnic background, and partisan background (an Obama appointee). In this context, we find evidence to suggest that some citizens may indeed have believed Curiel to be prejudiced because of his ethnicity.

Our results also unlock other questions concerning perceived ingroup favoritism, and specifically whether the effects that we observe are more pronounced among certain subgroups of the population. While we explore this question in a series of analyses in the appendix, we expect that there is much more to unpack in future work. If naked ingroup favoritism has an influence on assessments of legal impartiality, this would suggest graver implications than we have uncovered. We indicate, for the first time, one specific mechanism that may give rise to this pattern: citizen beliefs in the (in)ability of judges to engage in unbiased decision-making.

Works Cited

Allen, David, and Diane Wall. 1993. "Role Orientations and Women State Supreme Court
      Justices." *Judicature* 77(3): 156-165.

Anzia, Sarah, and Christopher Berry. 2011. "The Jackie (and Jill) Robinson Effect: Why Do
      Congresswomen Outperform Congressmen?" *American Journal of Political Science* 55(3): 478-
      493.

Baird, Vanessa, and Amy Gangl. 2006. "Shattering the Myth of Legality: The Impact of the
      Media's Framing of Supreme Court Procedures on Perceptions of Fairness." *Political
      Psychology* 27(4): 597-614.

Barreto, Matt, Mario Villarreal, and Nathan Woods. 2005. "Metropolitan Latino Political
      Behavior." *Journal of Urban Affairs* 27(1): 71-91.

Berinsky, Adam, Gregory Huber, and Gabriel Lenz. 2012. "Evaluating Online Labor
      Markets for Experimental Research." *Political Analysis* 20(3): 351-368.

Bertrand, Marianne, and Sendhil Mullainathan. 2004. "Are Emily and Greg More Employable than
      Lakisha and Jamal? An Experiment on Labor Market Discrimination." *American Economic
      Review* 94(4): 991-1013.

Boyd, Christina, Lee Epstein, and Andrew Martin. 2010. "Untangling the Causal Effects of Sex
      on Judging." *American Journal of Political Science* 54(2): 389-411.

Burden, Barry, Yoshikuni Ono, and Masahiro Yamada. 2017. "Reassessing Public Support for a
      Female President." *Journal of Politics* 79(3): 1073-1078.

Choi, Stephen, Mitu Gulati, Mirya Holman, and Eric A. Posner. 2011. "Judging Women." *Journal
      of Empirical Legal Studies* 8(3): 504-532.

Christenson, Dino, and David Glick. 2013. "Crowdsourcing Panel Studies and Real-Time
      Experiments in MTurk." *Political Methodologist* 20(2): 27-32.

Christenson, Dino, and David Glick. 2015. "Chief Justice Roberts's Health Care Decision

    Disrobed: The Microfoundations of the Supreme Court's Legitimacy." *American Journal of*

    Political Science 59(2): 403-418.

Devine, Patricia. 1989. "Stereotypes and Prejudice: Their Automatic and Controlled

    Components." *Journal of Personality and Social Psychology* 56(1): 5-18.

Dolan, Kathleen. 2010. "The Impact of Gender Stereotyped Evaluations on Support for Women

    Candidates." *Political Behavior* 32(1): 69-88.

Dolan, Kathleen. 2014. "Gender Stereotypes, Candidates Evaluations, and Voting for Women

    Candidates" *Political Research Quarterly* 67(1): 96-107.

Eagly, Alice, and Steven Karau. 2002. "Role Congruity Theory of Prejudice Toward Female

    Leaders." *Psychological Review* 109(3): 573-598.

Fahrang, Sean, and Gregory Wawro. 2004. "Institutional Dynamics on the U.S. Court of Appeals:

    Minority Representation under Panel Decision-Making." *Journal of Law, Economics and*

    Organization 20(2): 299-330.

Fridkin, Kim, and Patrick Kenney. 2009. "The Role of Gender Stereotypes in U.S. Senate

    Campaigns." *Politics & Gender* 5(3): 301-324.

Gaddis, S. Michael. 2017. "How Black are Lakisha and Jamal? Racial Perceptions from Names Used

    in Correspondence Audit Studies." *Sociological Science* 4: 469-489.

Gibson, James L. 2008. "Challenges to the Impartiality of State Supreme Courts: Legitimacy Theory

    and 'New Style' Judicial Campaigns." *American Political Science Review* 102(1): 59-75.

Gibson, James L., and Gregory Caldeira. 2009. "Knowing the Supreme Court? A Reconsideration of

    Public Ignorance of the High Court." *Journal of Politics* 71(2): 429-441.

Gibson, James L., and Gregory Caldeira. 2011. "Has Legal Realism Damaged the Legitimacy of the

    U.S. Supreme Court?" *Law & Society Review* 45(1): 195-219.

Gibson, James L., and Gregory Caldeira. 2012. "Campaign Support, Conflicts of Interest, and

      Judicial Impartiality: Can Recusals Rescue the Legitimacy of Courts?." *Journal of Politics* 74(1):

      18-34.

Gibson, James L., and Michael J. Nelson. 2017. "Reconsidering Positivity Theory: What Roles do

      Politicization, Ideological Disagreement, and Legal Realism Play in Shaping U.S. Supreme

      Court Legitimacy?" *Journal of Empirical Legal Studies* 14(3): 592-617.

Glynn, Adam, and Maya Sen. 2015. "Identifying Judicial Empathy: Does Having Daughters

      Cause Judges to Rule for Women's Issues." *American Hournal of Political Science* 59(1): 37-54.

Hainmueller, Jens, Daniel Hopkins, and Teppei Yamamoto. 2014. "Causal Inference in Conjoint

      Analysis: Understanding Multi-Dimensional Choices via Stated Preference Experiments."

      Political Analysis 22(1): 1-30.

Hainmueller, Jens, Dominik Hangartner, and Teppei Yamamoto. 2015. "Validating Vignette and

      Conjoint Survey Experiments against Real-world Behavior." *Proceedings of the National Academy*

      of Sciences of the United States of America 112(8): 2395-2400.

Hayes, Danny. 2011. "When Gender and Party Collide: Stereotyping in Candidate Trait

      Attribution." *Politics & Gender* 7(2): 133-165.

Hibbing, John., and Elizabeth Theiss-Morse. 2002. *Stealth Democracy: Americans' Beliefs about How*

      Government Should Work. Cambridge: Cambridge University Press.

Huddy, Leonie, and Nayda Terkildsen. 1993. "Gender Stereotypes and the Perception of Male and

      Female Candidates." *American Journal of Political Science* 37(1): 119-147.

Huddy, Leonie. 2001. "From Social to Political Identity." *Political Psychology* 22(1): 127-156.

Jardina, Ashley. 2019. *White Identity Politics.* Cambridge: Cambridge University Press.

Kaslovsky, Jaclyn, Jon Rogowski, and Andrew Stone. Forthcoming. "Descriptive

Representation and Public Support for Supreme Court Nominees." *Political Science Research and Methods*. doi:10/1017/psrm.2019.59.

Kastellac, Jonathan. 2016. "Race, Context, and Judging on the Courts of Appeals: Race-Based Panel Effects in Death Penalty Cases." Working paper. <http://www.princeton.edu/~jkastell/Death_Penalty_Race/kastellec_race_DP_paper.pdf>

Kastellac, Jonathan, Jeffrey Lax, and Justin Phillips. 2011. "Public Opinion and Senate Confirmation of Supreme Court Nominees." *Journal of Politics* 72(3): 767-784.

Kinder, Donald., and Nathan Kalmoe. 2017. *Neither Liberal nor Conservative: Ideological Innocence in the American Public*. Chicago: University of Chicago Press.

Koch, Jeffrey. 2000. "Do Citizens Apply Gender Stereotypes to Infer Candidates' Ideological Orientations?" *Journal of Politics* 62(2): 414-429.

Koch, Jeffrey. 2002. "Gender Stereotypes and Citizens' Impressions of House Candidates' Ideological Orientations." *American Journal of Political Science* 46(2): 453-462.

Lawless, Jennifer. 2004. "Women, War, and Winning Elections: Gender Stereotyping in the Post-September 11th Era." *Political Research Quarterly* 57(3): 479-490.

Lawrence, Regina, and Melody Rose. 2014. "The Race for the Presidency: Hillary Rodham Clinton." In *Women and Elective Office*, eds. Sue Thomas and Clyde Wilcox. New York: Oxford University Press, 67–79.

Lupia, Arthur. 1994. "Shortcuts versus Encyclopedias: Information and Voting Behavior in California Insurance Reform Elections." *American Political Science Review* 88(1): 63-76.

Lupia, Arthur. 2016. *Uninformed: Why People Know so Little about Politics and What We Can Do about It*. New York: Oxford University Press.

Mason, Lilliana. 2015. "'I Respectfully Disagree': The Differential Effects of Partisan Sorting on Social and Issue Polarization." *American Journal of Political Science* 59(1): 128-145.

Mason, Lilliana, and Julie Wronski. 2018 "One Tribe to Bind Them All: How Our Social Group Attachments Strengthen Partisanship." *Political Psychology* 39(1): 257-277.

Matson, Marsha, and Terri Susan Fine. 2006. "Gender, Ethnicity, and Ballot Information." *State Politics & Policy Quarterly* 6(1): 49-72.

McDermott, Monika. 1997. "Voting Cues in Low-Information Elections: Candidate Gender as a Social Information Variable in Contemporary United States Elections." *American Journal of Political Science* 41(1): 270-283.

McDermott, Monika. 1998. "Race and Gender Cues in Low-Information Elections." *Political Research Quarterly* 51(4): 895-918.

Meyer, Chase, and J. David Woodard. 2017. "It's Not Race, It's Politics! A Natural Experiment Examining the Effects of Race in Electoral Politics." *Social Science Quarterly* 98(1): 120-131.

Miller, Susan, and Shana Maier. 2008. "Moving Beyond Numbers: What Female Judges Say about Different Judicial Voices." *Journal of Women, Politics & Policy* 29(4): 527-559.

Nelson, Kjersten. 2015. "Double-Bind on the Bench: Citizens' Perceptions of Judge Gender and the Court." *Politics & Gender* 11(2): 235-264.

Reddick, Malia, Michael Nelson, and Rachel Paine Caulfield. 2009. "Racial and Gender Diversity on State Courts." *Judges' Journal* 48(3): 28-32.

Rose, Melody. 2013. *Women & Executive Office*. Boulder: Lynne Rienner Publishers.

Sanbonmatsu, Kira. 2002. "Gender Stereotypes and Vote Choice." *American Journal of Political Science* 46(1): 20-34.

Scherer, Nancy, and Brett Curry. 2010. "Does Descriptive Race Representation Enhance Institutional Legitimacy?" *Journal of Politics* 72(1): 90-104.

Sen, Maya. 2015. "Is Justice Really Blind? Race and Reversal in U.S. Courts." *Journal of Legal Studies* 44(S1): S187-S229.

Sen, Maya. 2017. "How Political Signals Affect Public Support for Judicial Nominations: Evidence from a Conjoint Experiment." *Political Research Quarterly* 70(2): 374-393.

Sides, John, Michael Tesler, and Lynn Vavreck. 2018. *Identity Crisis: The 2016 Presidential Campaign and the Battle for the Meaning of America*. Princeton: Princeton University Press.

Tesler, Michael. 2012. "The Return of Old-Fashioned Racism to White Americans Partisan Preferences in the Early Obama Era." *Journal of Politics* 75(1): 110-123.

Tokeshi, Matthew, and Tali Mendelberg. 2015. "Countering Implicit Appeals: Which Strategies Work?" *Political Communication* 32(4): 648-672.

Tyler, Tom. 1988. "What is Procedural Justice? Criteria Used by Citizens to Assess the Fairness of Legal Procedures." *Law & Society Review* 22(1): 103-136.

Tyler, Tom. 1990. *Why People Obey the Law*. Princeton: Princeton University Press.

Tyler, Tom. 2006. "Psychological Perspectives on Legitimacy and Legitimation." *Annual Review of Psychology* 57: 375-400.

Tyler, Tom, and Yuen Huo. 2002. *Trust in the Law: Encouraging Public Cooperation with the Police and Courts*. New York: Russell Sage Foundation.

Tyler, Tom. 2005. "Policing in Black and White: Ethnic Group Differences in Trust and Confidence in the Police." *Police Quarterly* 8(3): 322-342.

Tyler, Tom. 2006. "Psychological Perspectives on Legitimacy and Legitimation." *Annual Review of Psychology* 57: 375-400.

Tyler, Tom, and Yuen Huo. 2002. *Trust in the Law: Encouraging Public Cooperation with the Police and Courts*. New York: Russell Sage Foundation.

Valentino, Nicholas, and Vincent Hutchings. 2004. "The Centrality of Race in American Politics." *Annual Review Political Science* 7: 383-408.

Van Hall, Karey. 2017. "Trump's Attacks on Federal Judge Could Backfire." *Politico*.

<https://www.politico.com/story/2017/02/trump-judge-attack-backfire-234649>

Vining, Richard, and Teena Wilhelm. 2011. "Measuring Case Salience in State Courts of Last

Resort." *Political Research Quarterly* 64(3): 559-572.

Supplemental Information for Study 1

Experimental Stimuli

Case type	Vignette
Abortion (civil)	**Abortion Clinic Case Assigned to Judge [NAME]** March 15, 2018 WASHINGTON (AP) — The fate of a proposed abortion clinic now rests in the hands of a federal court after Citizens for a Pro-Life Society filed suit to prevent its opening. Judge [NAME] will hear arguments in the case this fall. Observers say that his ruling could further inflame a simmering debate about abortion rights. At issue is a Planned Parenthood clinic that is slated to open next year in Auburn Hills. When plans for the site were unveiled, pro-life groups complained that they violated a zoning ordinance that restricts the areas in which abortion providers can operate. "We demand that it be shut down," said Mark Milner, the society's director. "The activity is not only wrong, but unlawful in this community." Now, [NAME], the judge in the case, must decide whether Planned Parenthood followed proper licensing procedures and, more generally, whether the licensing requirements are overly burdensome on abortion providers. The 48-year old Judge [NAME], a 2-year veteran of the bench, will be facing [his/her] first major abortion case. The implications of the ruling for Auburn Hills are substantial. The nearest clinic outside of the proposed Planned Parenthood site sits some 135 miles away.
Immigration (criminal)	**Immigrant's Assault Case Assigned to Judge [NAME]** March 15, 2018 WASHINGTON (AP) — The fate of an unlawful immigrant charged with a violent assault now rests in the hands of a federal court. Judge [NAME] will hear arguments in the case this fall. Observers say that his ruling could further inflame a simmering debate about immigration. At issue in the case is an altercation that began in an Auburn Hills bar but then spilled outside. The accused, an immigrant from El Bajio, allegedly pulled a box cutter to attack another man, resulting in severe injuries to his head and chest. "The government is really pursuing immigrant-caused crimes very aggressively these days," said Ken Montgomery, an analyst for the Legal Research Initiative. "These efforts raise questions about whether different classes of individuals are being treated differently for the same actions." Now, [NAME], the judge in the case, must evaluate the evidence and, more generally, the government's responsibilities when prosecuting charges against illegal immigrants. The 48-year old Judge [NAME], a 2-year veteran of the bench, will be facing [his/her] first major immigration case. The implications of the ruling for the government's ability to prosecute are substantial. At least three similar cases are on the judiciary's docket in the coming months.

Dependent variable battery (factor loadings in parentheses)
- Judge [name] will display improper bias when ruling on the [abortion/immigration] case (0.78)
- Judge [name]'s values and political views will influence the ruling in the [abortion/immigration] case (0.64)
- Judge [name] should be required to recuse [himself/herself] from the case (0.76)

Manipulation check
- What race/ethnicity do you think Judge [name] is likely to be?

Sample Characteristics

Characteristic	Proportion of Sample
Age	
18-25	11.07
26-35	18.93
36-45	16.88
46-55	16.07
56-65	19.47
66 or over	17.58
Race	
White	70.52
Black	11.26
Asian/Pacific Isl.	5.26
Native American	0.96
Other/Refused	12.00
Ethnicity	
Hispanic	13.67
Non-Hispanic	86.33
Gender	
Male	46.97
Female	53.03

Dependent Variable
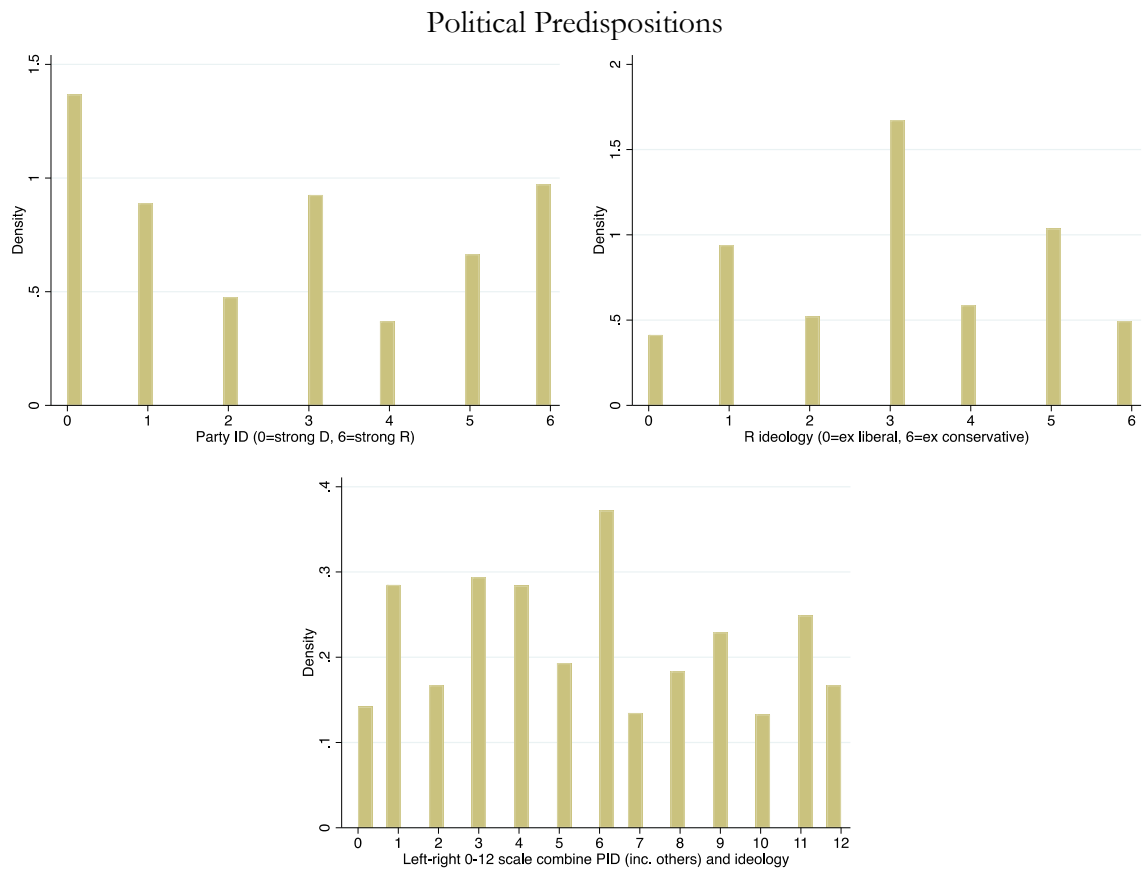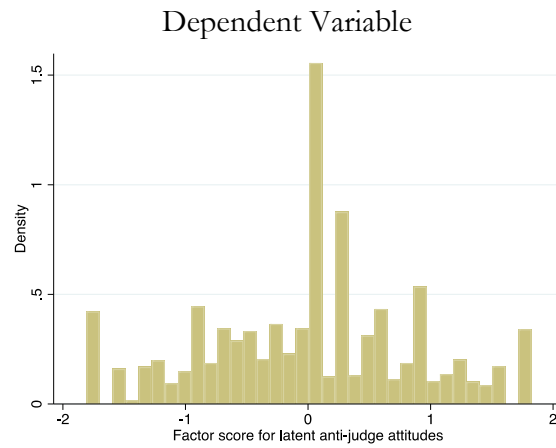
Political Predispositions

Study 1, Part 1: Non-interactive specification

	Pooled Results
Female judge condition	-0.07*
	(0.03)
Hispanic judge condition	-0.15*
	(0.04)
African American judge condition	-0.09*
	(0.04)
Immigration case condition	-0.15*
	(0.03)
Left-right predispositions	0.01*
	(0.004)
Female	-0.11*
	(0.03)
Black	0.13*
	(0.05)
Hispanic	0.15*
	(0.04)
Education	-0.07*
	(0.03)
Knowledge	-0.14*
	(0.02)
R^2	0.050
N	3117

Note: Dependent variable is the factor score that combines the bias, recusal, and values items. Model uses an OLS specification, with standard errors in parentheses. *p<0.05. Baseline condition is a white male judge.

Study 1, Part 2: Analysis of Separated DVs, Including Judge Bias Item Only

	Pooled		Abortion only		Immigration only	
	Bias	Recusal	Bias	Recusal	Bias	Recusal
Female judge	-0.21**	-0.24**	-0.11	-0.21*	-0.31**	-0.27*
	(0.07)	(0.08)	(0.10)	(0.11)	(0.11)	(0.11)
Female judge	0.03**	0.03**	0.02	0.03*	0.03*	0.03
*L-R identity	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
Hispanic judge	-0.42**	-0.40**	-0.28*	-0.23	-0.55**	-0.57**
	(0.09)	(0.09)	(0.12)	(0.13)	(0.13)	(0.14)
Hispanic judge	0.04**	0.05**	0.01	0.01	0.06**	0.09**
* L-R identity	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
L-R identity	-0.01	-0.01	-0.0001	-0.01	-0.01	-0.02
	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
R^2	0.04	0.05	0.03	0.04	0.06	0.06
N	3117	3117	1568	1568	1549	1549

Note: Models use an OLS specification, with standard errors in parentheses. *p<0.05, **p<0.01. Baseline condition is a white male judge. Models also control for assignment to case condition, to the African American judge condition and its interaction with predispositions and a respondent's gender, race, ethnicity, education, and knowledge.

(a) Gender

(b) Race
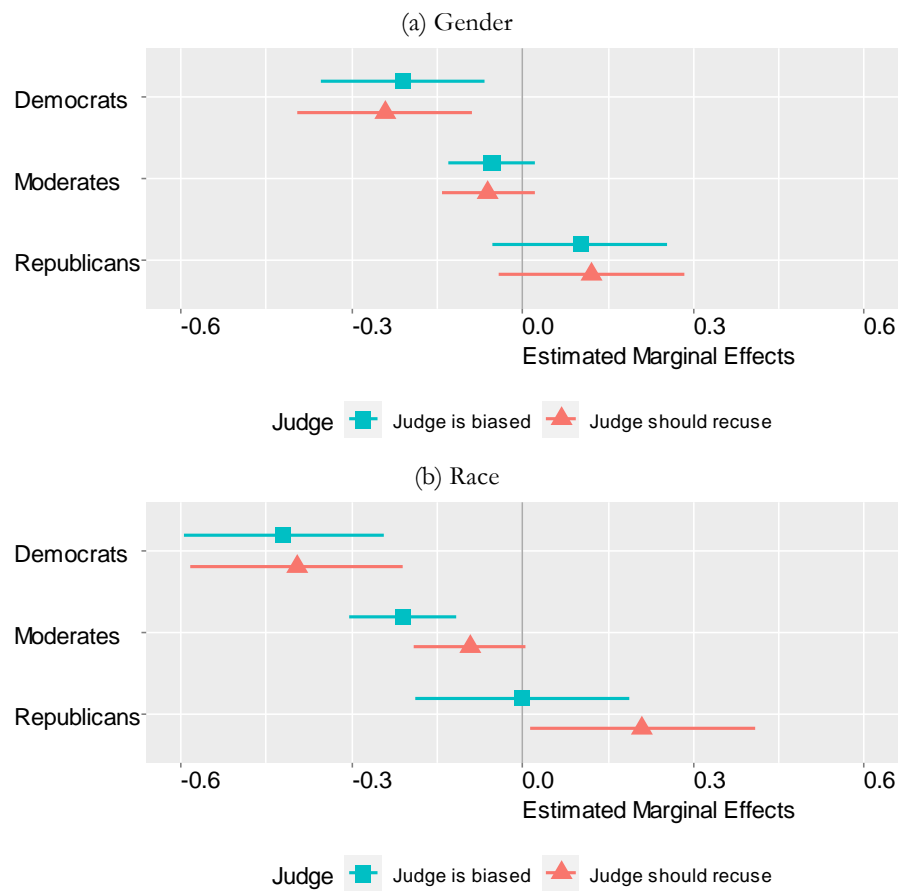
Note: Figures show the estimated marginal effects of judge gender and race on bias and recusal items from the models above.

Study 1, Part 3: Effect of Treatment on the Treated Analysis

	Pooled Results	Separated Results	
		Abortion case	Immigration case
Female judge condition	-0.15*	-0.14	-0.16
	(0.06)	(0.09)	(0.09)
Female judge condition *Left-right predispositions	0.02*	0.02	0.02
	(0.01)	(0.01)	(0.01)
Hispanic judge condition	-0.44**	-0.32**	-0.56**
	(0.07)	(0.10)	(0.10)
Hispanic judge condition *Left-right predispositions	0.04**	0.01	0.07**
	(0.01)	(0.02)	(0.02)
Left-right predispositions	-0.01	-0.003	-0.02
	(0.01)	(0.01)	(0.01)
R^2	0.05	0.05	0.06
N	2430	1214	1216

Note: Dependent variable is the factor score that combines the bias, recusal, and values items. To estimate the effect on the treated, respondents who failed the manipulation check have been removed. Models use an OLS specification, with standard errors in parentheses. *p<0.05, **p<0.01. Baseline condition is a white male judge. Models also control for assignment to case condition, to the African American judge condition and its interaction with predispositions and a respondent's gender, race, ethnicity, education, and knowledge.

Study 1, Part 4: Further Analysis of Separated Results

Perceptions of Gender Bias in Abortion Cases
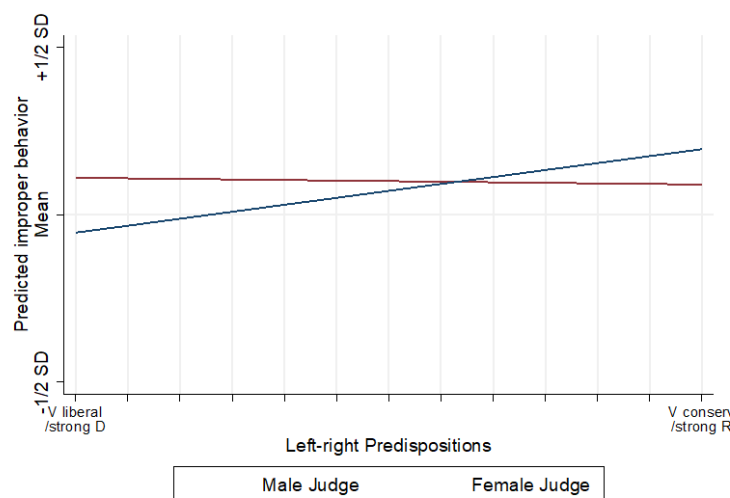
Note: Based on estimates from the main abortion model.

Perceptions of Hispanic Bias in Immigration Cases
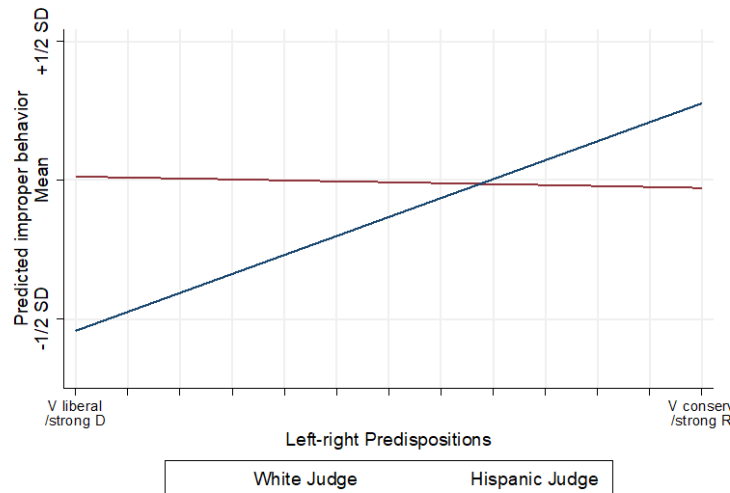
Note: Based on estimates from the main immigration model.

Study 1, Part 5: Intersectionality Analysis of Effects for Specific Judge Types (White Male, White Female, Hispanic Male, Hispanic Female)

Perceptions of Bias Based on Judge Type
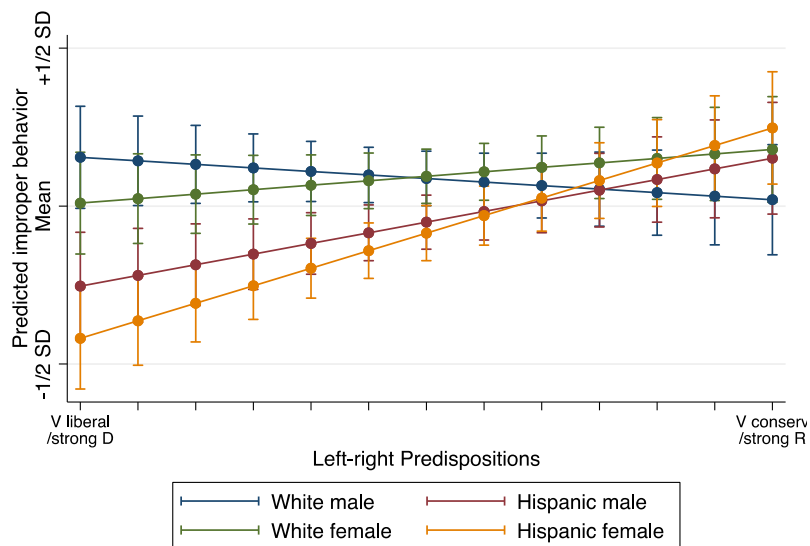
Note: Based on estimates from a model with dummies for judge types listed in the legend. Substantively, the negative slope on white male judges indicates that conservative Republicans perceive these judges as slightly *less* biased than do liberal Democrats. On the other hand, the slopes are positive and *increasing* in magnitude for white female, Hispanic male, and Hispanic female judges. In the most extreme case, the strong positive slope on Hispanic females indicates that conservative Republicans perceive these judges as much more biased than do liberal Democrats.

Study 1, Part 6: Analysis of How Respondent Gender Affects Perceptions of Judge Bias

How does respondent gender condition the effects of judge gender on perceived bias?
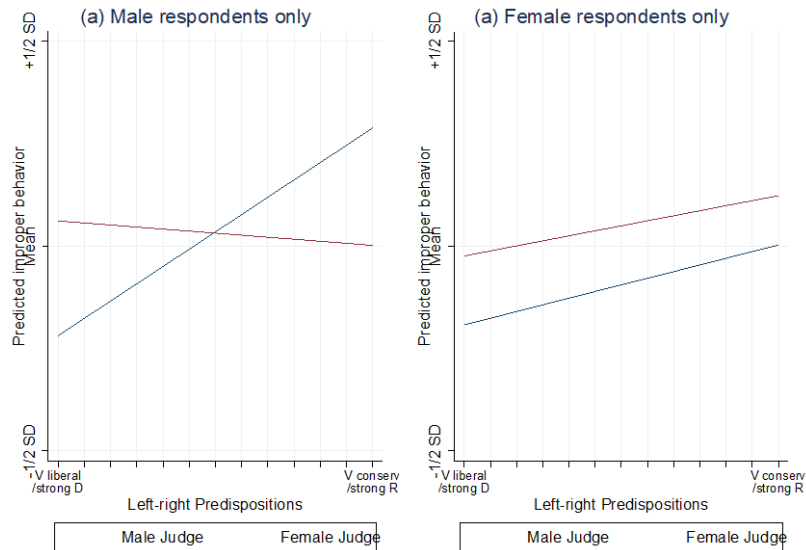
Note: Based on estimates from the main model run separately for male and female respondents. In substantive terms, these results provide insight about how identity politics has unique effects conditional on both the gender of the *judge* and of the *respondent*. Female respondents of all political persuasions see women judges as a bit less biased than their male counterparts. However, for male respondents, female judges are seen as very biased, but only among members of the political right. This provides further evidence of how perceptions of bias are conditional, but only for female judges, and it also suggests a role for in-group / out-group calculations in the judgment phase.

Study 1, Part 7: Difference-in-Means Test

	Left (Lib Dems)	Right (Cons Reps)	T-test
Female judge impropriety	-0.10 (0.03)	0.01 (0.04)	p=0.007*
Male judge impropriety	0.02 (0.03)	-0.01 (0.04)	p=0.233
Hispanic judge impropriety	-0.20 (0.04)	0.01 (0.04)	p=0.0004*
White judge impropriety	0.10 (0.03)	0.02 (0.05)	p=0.088

Note: Means and standard errors for perceived impropriety are displayed in the table. The significant differences indicate that liberal Democrats (conservative Republicans) are less (more) likely to perceive female judges to evince improper behavior. They also indicate that liberal Democrats (conservative Republicans) are less (more) likely to perceive Hispanic judges to evince improper behavior.

Supplemental Information for Study 2

Experimental Stimuli

Case type	Description
Abortion (civil)	A group of pro-life advocates has filed suit in order to prevent a new abortion clinic from opening. The lawsuit alleges that the proposed site of the clinic would violate local and state zoning ordinances that restrict the areas in which abortion providers can operate. Given the case's significant policy implications, legal experts argue that it could eventually end up in the federal court system on appeal.
Immigration (criminal)	A local prosecutor announced aggravated assault charges against an illegal immigrant following a knife attack that left another man with injuries to his head and chest. The assault is alleged to have occurred after a fight that began in a local bar spilled outside. Given the government's increasingly aggressive prosecutions of crimes committed by immigrants, the case has significant policy implications and legal experts argue that it could eventually end up in the federal court system on appeal.

Question Wording

After reading judge profiles …

- Which of the two judges do you think is more likely to have their values and political views influence how the decide the [abortion/immigration] case?
- Which of the two judges do you think is more likely to display improper bias when ruling on the case?

Sample Characteristics

Characteristic	Proportion of Sample
Age	
18-25	10.03
26-35	37.83
36-45	23.83
46-55	15.02
56-65	9.73
66 or over	3.56
Race	
White	82.64
Black	9.09
Asian/Pacific Isl.	6.29
Native American	0.5
Other	1.48
Ethnicity	
Hispanic	12.71
Non-Hispanic	87.29
Gender	
Male	44.75
Female	55.25

Study 2, Part 1: Pooled Conjoint Results by Respondent's Partisanship (Figure 2)

Average Component Interaction Effects (personal values influence decisions)
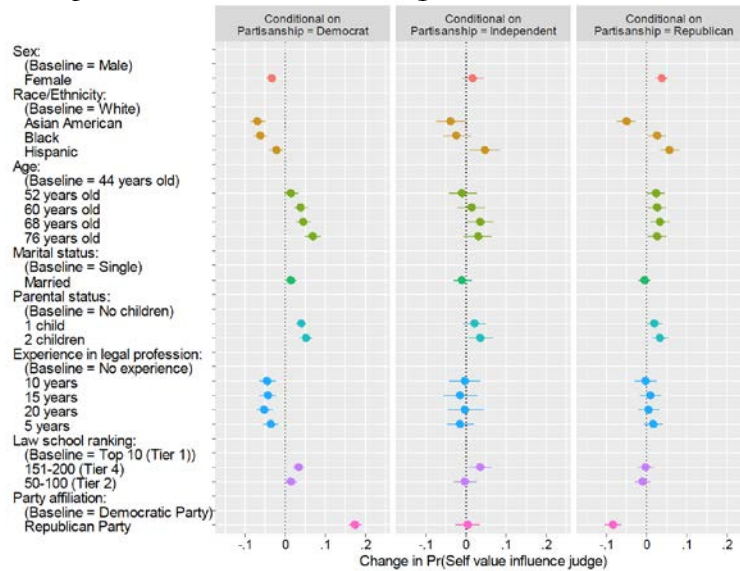
Study 2, Part 2: Separated Conjoint Results for Abortion Case (Figure 3)

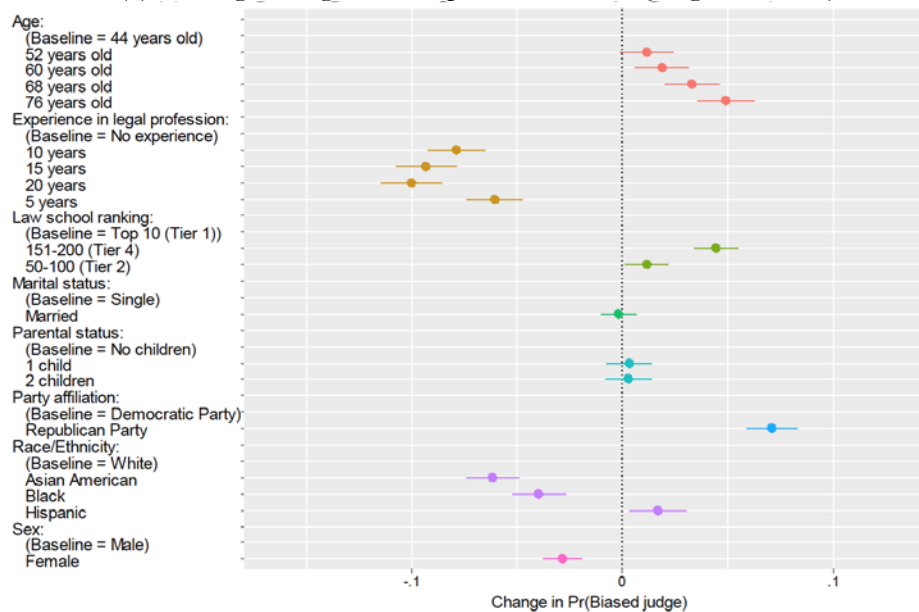Average Component Interactive Effects (personal values influence decisions)
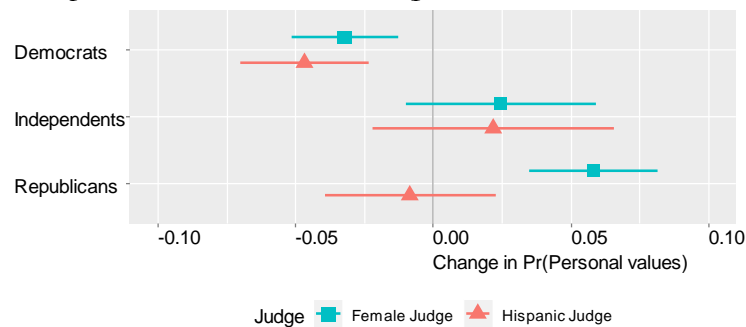
Study 2, Part 3: Separated Conjoint Results for Immigration Case (Figure 4)

Average Component Interactive Effects (personal values influence decisions)
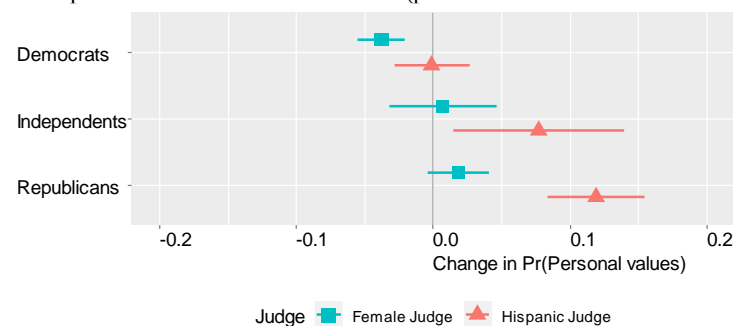
Study 2, Part 4: Pooled AMCEs

(a) Average Marginal Component Effects (personal values influence decisions)
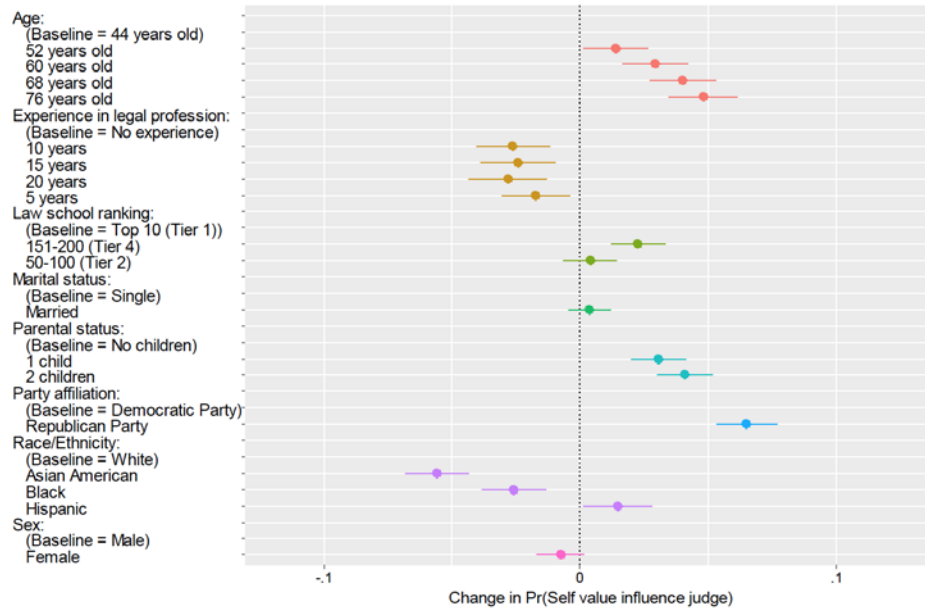
(b) Average Marginal Component Effects (judge is biased)

Study 2, Part 5: Separated AMCEs – Abortion Case

(a) Average Marginal Component Effects in Abortion Case (personal values influence decisions)
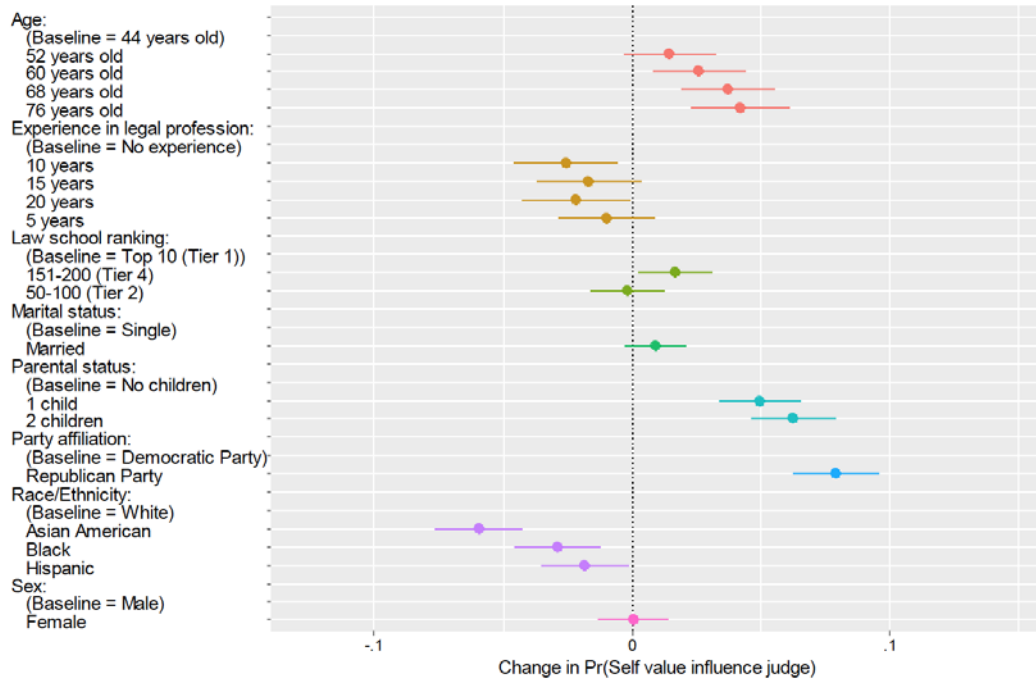
(b) Average Marginal Component Effects in Abortion Case (judge is biased)
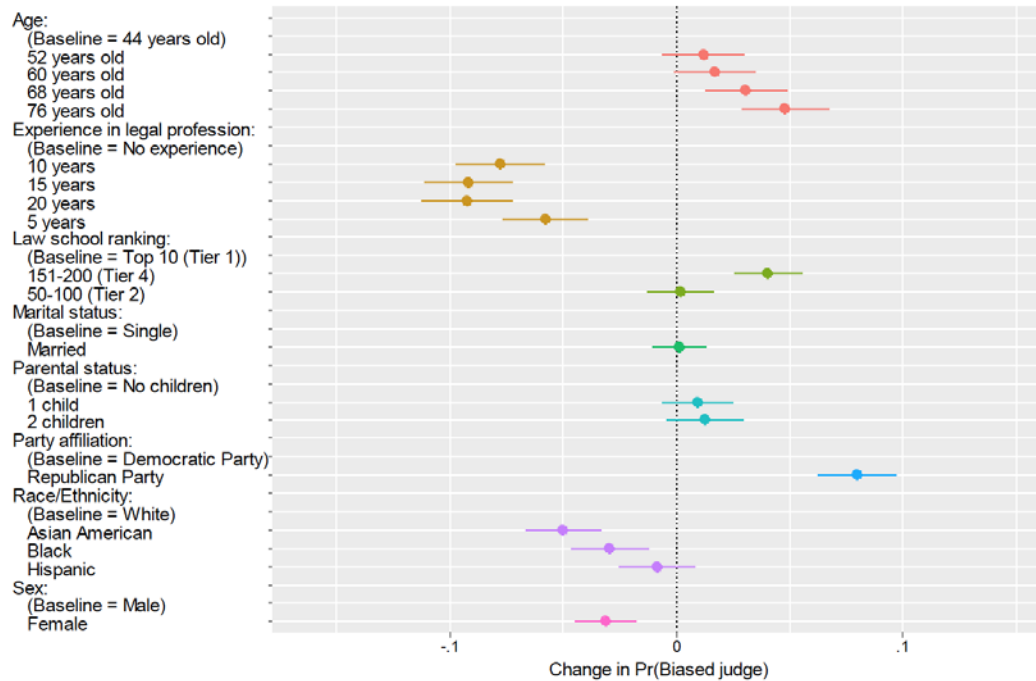
## Study 2, Part 6: Separated AMCEs – Immigration Case

(a) Average Marginal Component Effects in Immigration Case (personal values influence decisions)
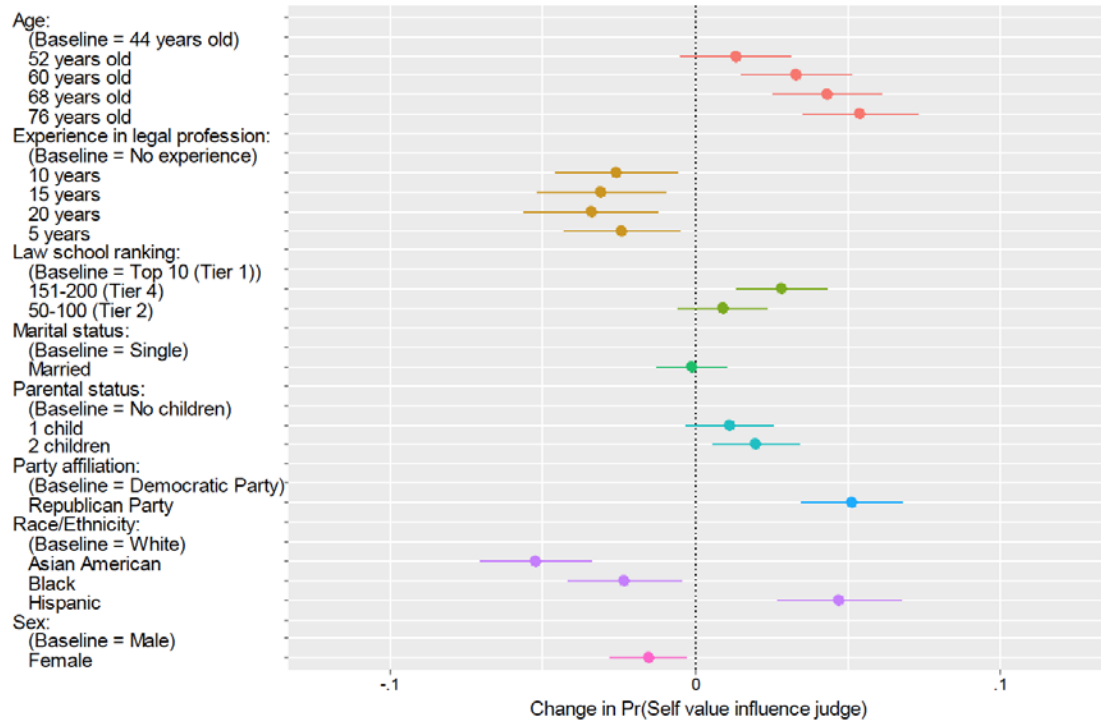
(b) Average Marginal Component Effects in Immigration Case (judge is biased)
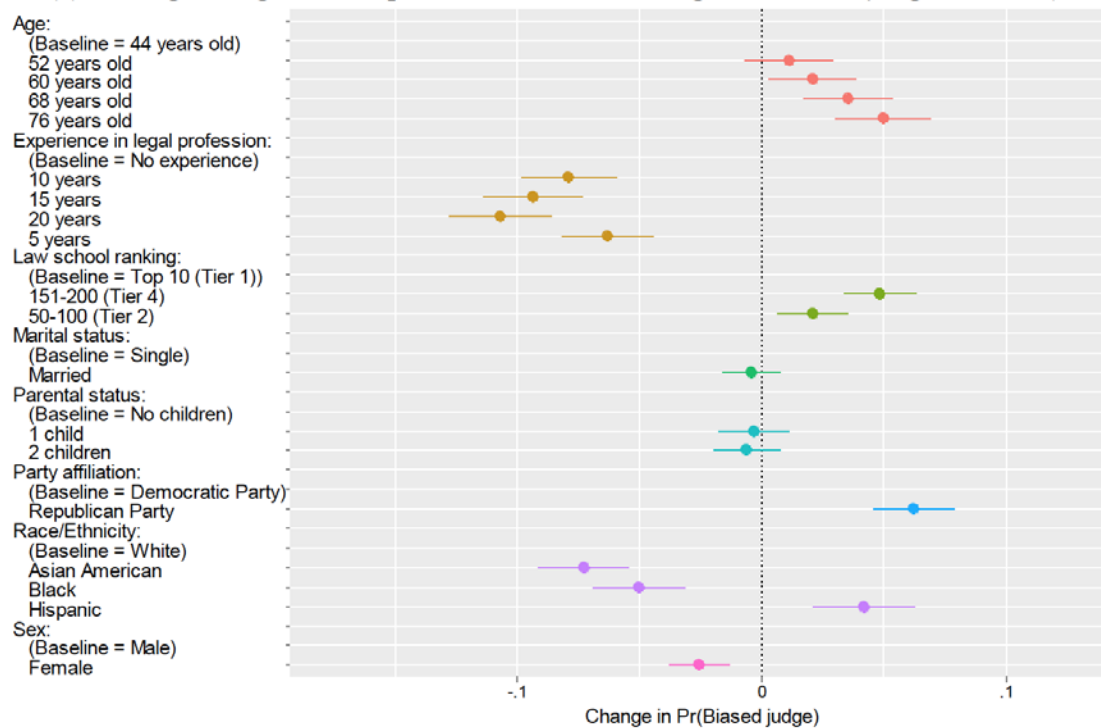
Study 2, Part 7: Full ACIEs for Figure 3

(a) Average Component Interactive Effects in Abortion Case (personal values influence decisions)
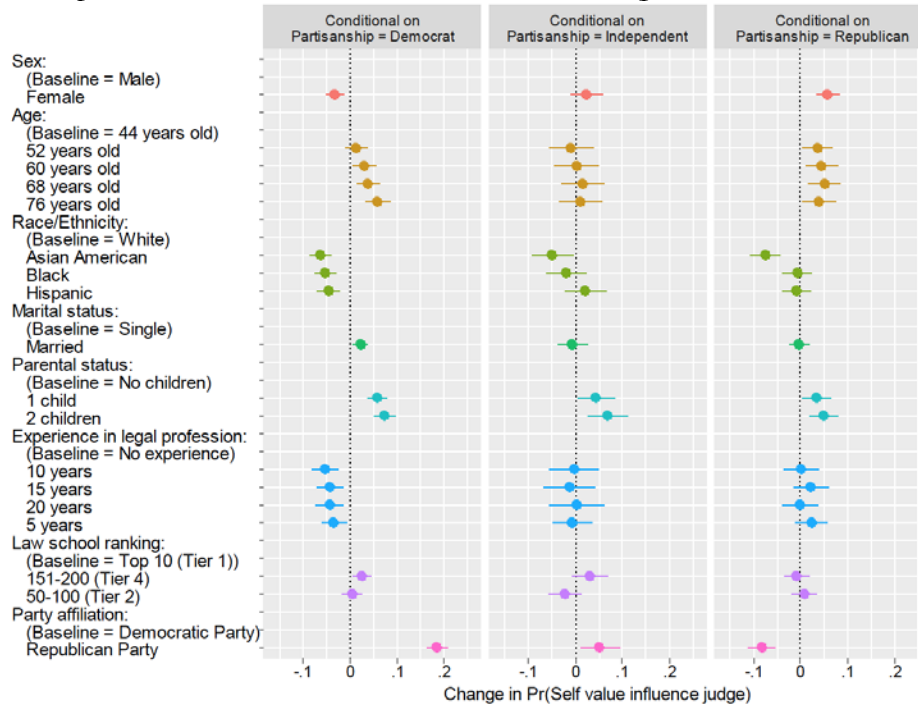

(b) Average Component Interactive Effects in Abortion Case (judge is biased)
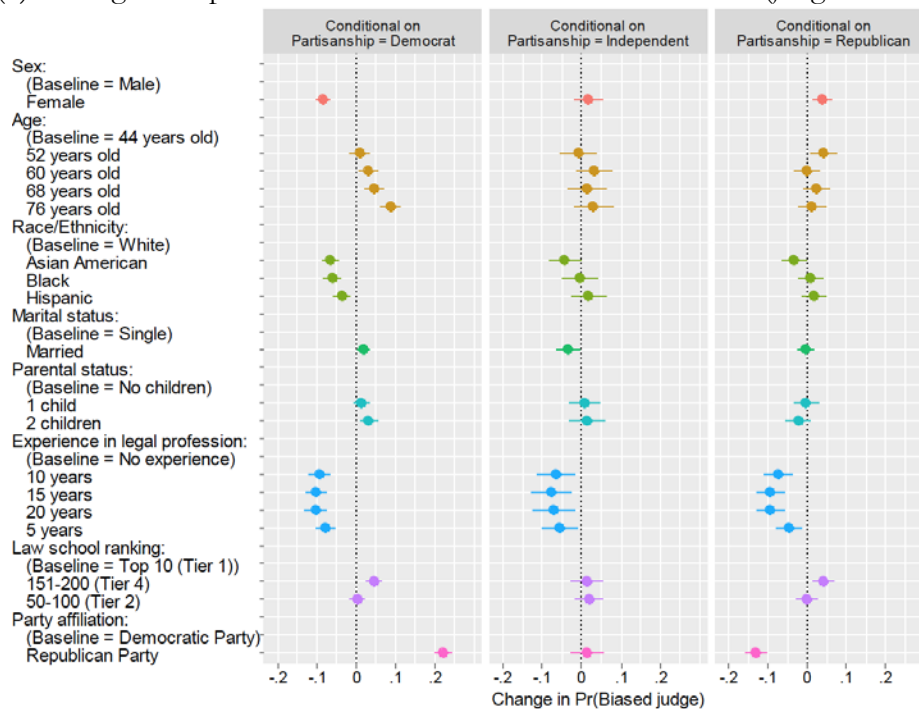
Study 2, Part 8: Full ACIEs for Figure 4

(a) Average Component Interactive Effects in Immigration Case (personal values influence decisions)
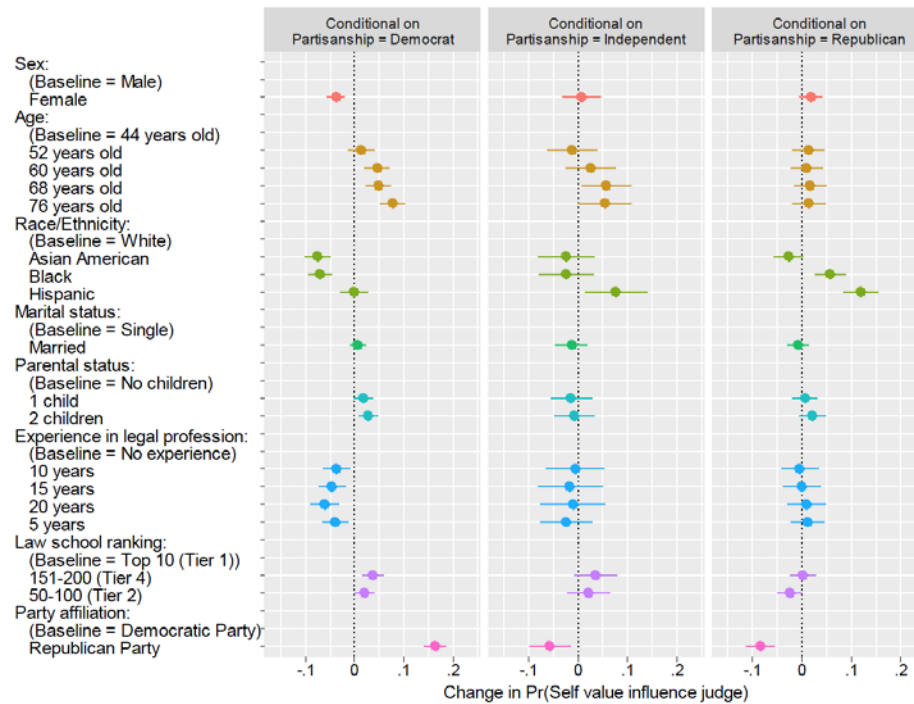
(b) Average Component Interactive Effects in Immigration Case (judge is biased)
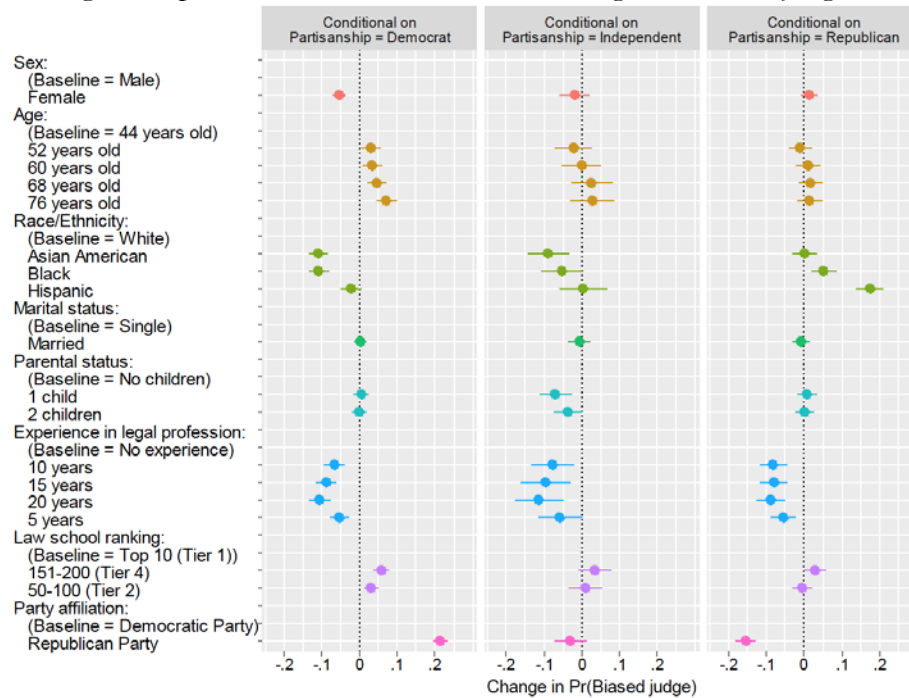
Study 2, Part 9: Comparison between a Republican Judge and a Democratic Judge

In the conjoint analysis, we have randomized judge partisanship in addition to judge race and gender. This means that our results for race and gender are displaying their average effects after controlling for the effects of judge partisanship; in other words, conditional on a respondent knowing a judge's party affiliation, what are the effects of the judge's race and gender? To add nuance to our interpretations, we display below the ACIEs of judges' attributes (including race and gender) conditional on the judge's party affiliation, which indicate how the effects of judge race and gender may differ by judge party.

(a) ACIE (personal values influence decisions)   (b) ACIE (judge is biased)
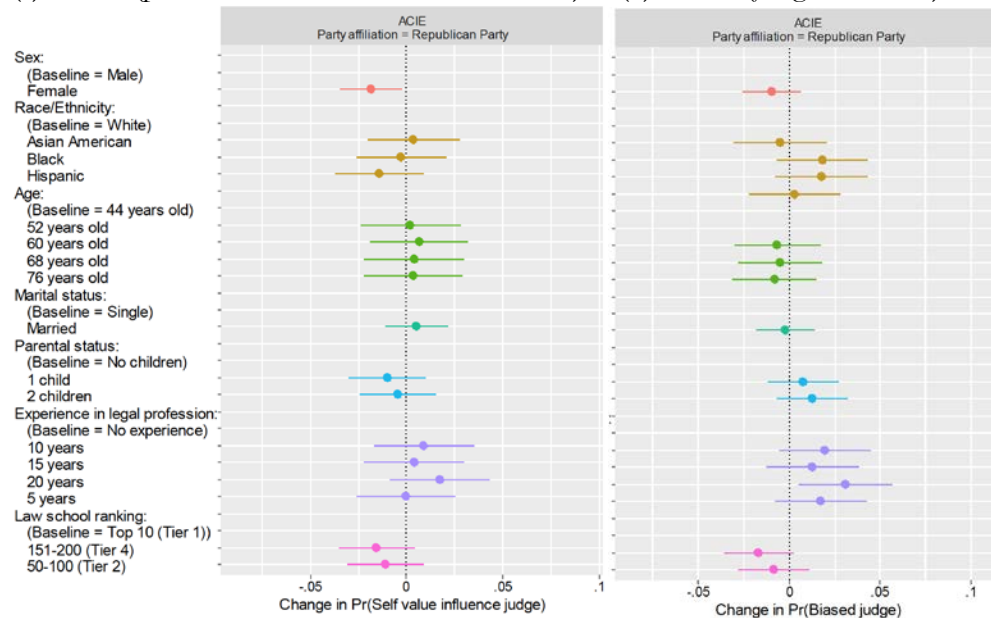
The results show that the effects of judge's race and gender do not vary depending on the judge's party affiliation. Respondents evaluate judges almost equally regardless of whether those judges are Democrats or Republicans. The only exception is that the effects of judge gender on the values item differs between a Republican judge and a Democratic judge: respondents perceive that Republican female judges are less likely to reflect their values and political views in their rulings when compared with Republican male judges with the identical attributes. Overall, the insignificant effects presented above demonstrate that a judge's partisan policy preferences have little effect on how their race and gender shape perceptions.

This pattern is more clearly shown in the figures below that display the ACIEs of judges' attributes conditional on the context of comparison, which directly captures whether a respondent and judge are "intraparty" (they share the same partisan identification) or "interparty" (their party affiliations are distinct). The effects of judge attributes are similar overall regardless of whether the pairings of judges are affiliated with the opposing parties (interparty condition) or with the same party (intraparty condition). Again, one exception is the effect of judge gender when asked about whether personal values influence judges' rulings. Similar to the results shown in the above figures, respondents perceive that female judges are less likely to reflect their own values and political views to their rulings than male judges only when both judges are affiliated with the Republican Party.

51

(a) Average Component Interactive Effects (personal values influence decisions)
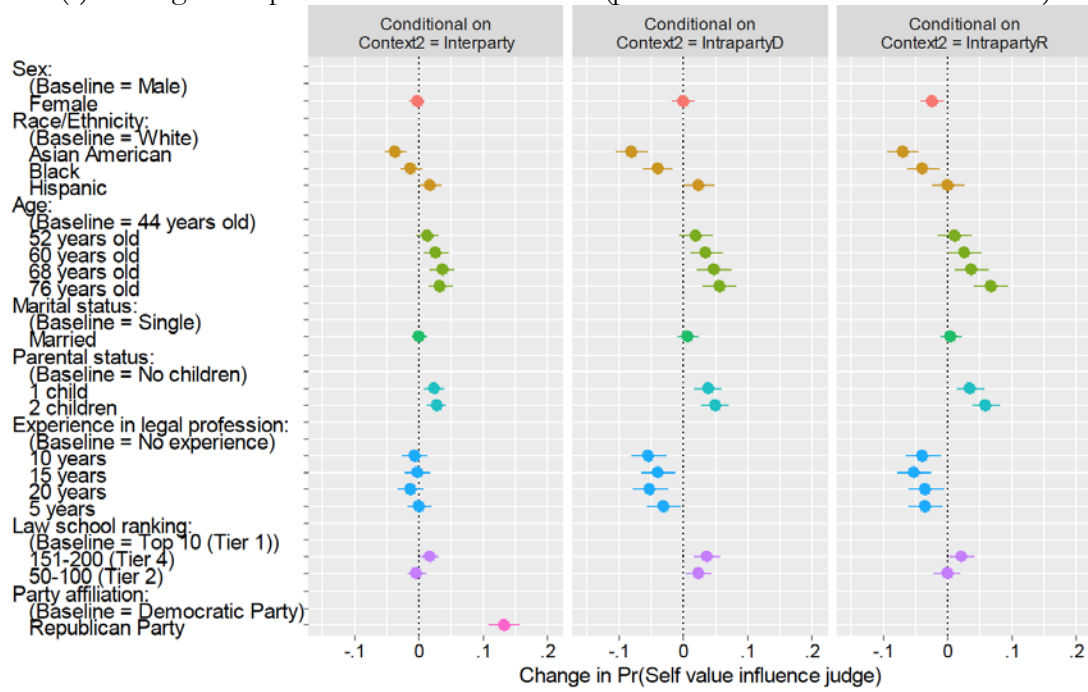
(b) Average Component Interactive Effects (judge is biased)
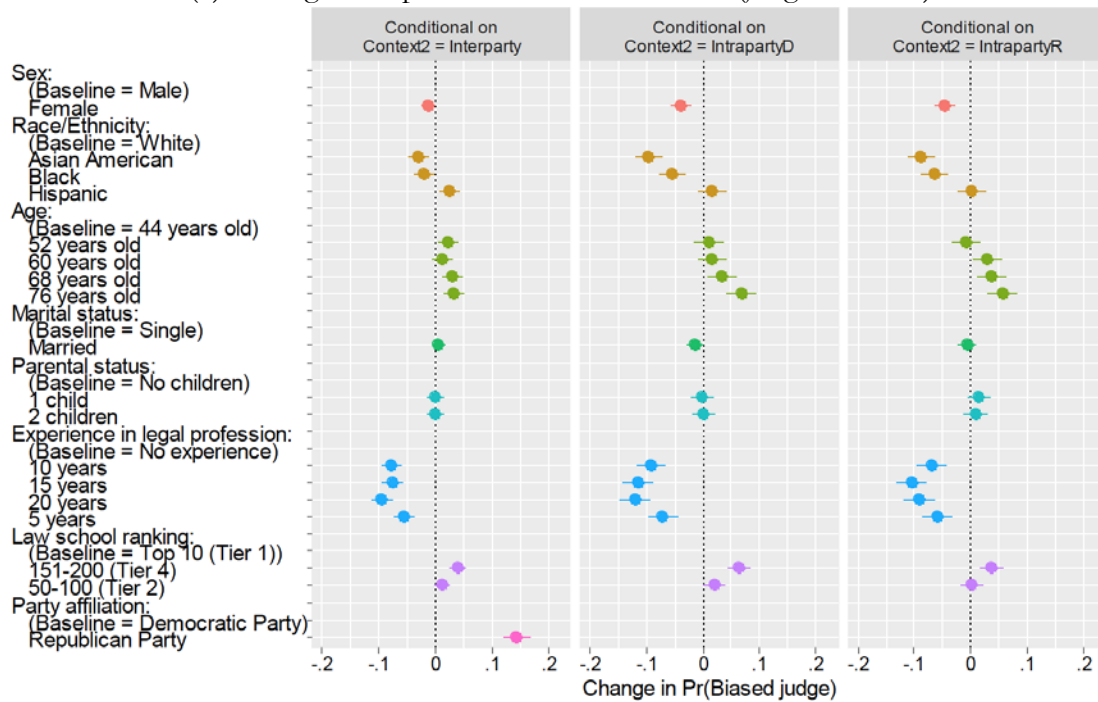
Note: The left panel shows parings of Democratic and Republican judges, the middle panel shows parings of only Democratic judges, and the right panel shows pairings of only Republican judges.

Study 2, Part 10: Gender affinity effects

The figures below display the ACIEs of judges' attributes (including race and gender) conditional on respondent gender, which indicates how the effects of judge race and gender may differ between male and female respondents. The results show that female respondents consider female judges to be less biased than male judges, while male respondents evaluate male and female judges equally. It appears that there exists gender affinity effect only among female respondents. This outcome may be partly because male respondents exhibit different attitudes toward female judges depending on their partisanship; i.e., male Republican respondents evaluate female judges differently from male Democratic respondents (see study 2, part 6 appendix).

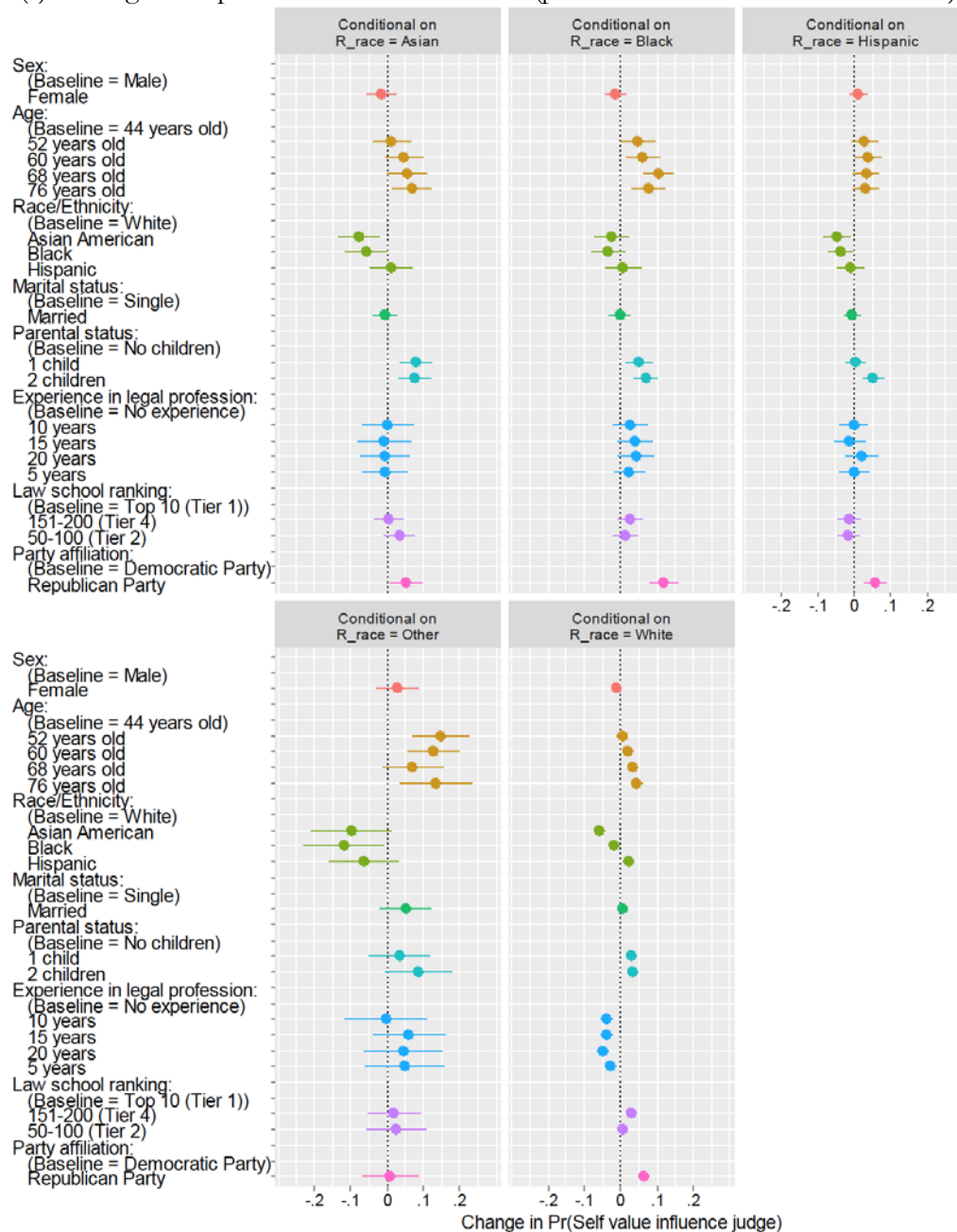(a) ACIE (personal values influence decisions)   (b) ACIE (judge is biased)
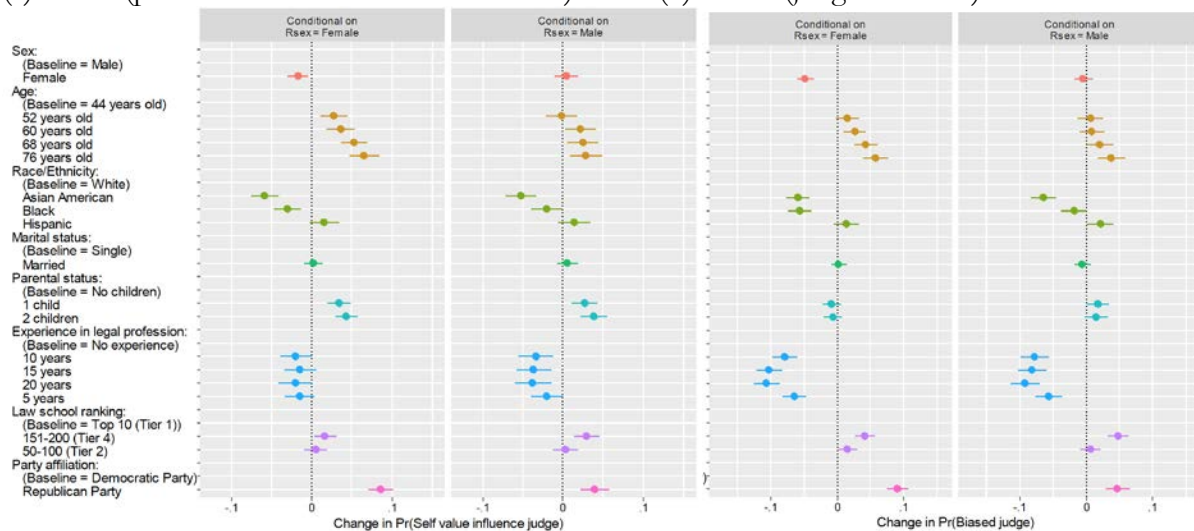
Study 2, Part 11: Race affinity effects

The next figures show the ACIEs of judges' attributes (including race and gender) conditional on respondent's race. The results show the effects of judge's race differ depending on respondent's race. Interestingly, Hispanic respondents evaluate Hispanic judges in the same terms as White judges (i.e., there is no race affinity effect among Hispanic respondents); in contrast, White respondents consider Hispanic judges to be more biased than White judges.

(a) Average Component Interactive Effects (personal values influence decisions)

(b) Average Component Interactive Effects (judge is biased)