Final Project: Subjectivity of judicial proceedings

Carlos Rafael Garduño Acolt

CS112 Fall 2020

Decision Memo for a Senator / Policy Maker

Executive Summary

The present decision memo discusses the issue of the subjectivity of judicial proceedings. More specifically, how it is that the identities of judges affect the outcomes of court applications. The analysis of the issue describes the context of the problem, and leads to the question: How might we measure how the identity of a judge affects the outcome of judicial proceedings? The methods section describes the techniques and results used in a relevant study regarding the differences between the conclusions female and male judges get to, while evaluating claims filled by men and women. The action plan goes a step farther and describes a partial replication of the paper, discussing its results, and how they fit with the original paper's, supporting the need for profound jurisdictional reform.

Analysis of the issue

Judicial proceedings are supposed to be fair and impartial. Decisions taken by one judge for a particular case should be similar to the decisions of a different judge handling a similar case. However, this rarely holds true as judges are fallible too. The identity, knowledge, and past experiences of a judge can most certainly affect the outcome of their proceedings. This widespread concern characterizes the current state of the problem. Every year, thousands of people are charged unfairly around the world, making the goal state clear, a reformed court system that is objective and unbiased. Although the main cause of the problem relies on the requirement of a human node between people and the justice system, it would be very hard, if not impossible, to imagine otherwise. We must take into account that cases are highly complex and nuanced, requiring human understanding to lead to reasonable and fair conclusions (ideally).

We then ask the question, how might we measure how the identity of a judge affects the outcome of judicial proceedings?

Methods

A prime piece of evidence of such subjectivity is discussed by Voeten, E. (2020) in the paper "Gender and judging: evidence from the European Court of human rights." In the study, the author answers the question: "How does increasing the number of women on the bench affect the outcome of judicial proceedings?" And at the heart of the analysis, investigating whether male and female judges reach different conclusions while evaluating claims filled by men and women. The study design is centered around two datasets describing cases and judges from the European Court of Human Rights. The first set incorporates the gender and kind of issue filled by applicants to the court. The second involves the gender and other identity traits of judges on the bench of the court, who were responsible for hearing the issues described in the first data set. Those identity traits include, among others, whether the judge identifies as part of the left or right of the chamber, whether they identify as religious or not, and whether they identify as nationalists or not. These variables often referred to as covariates, represent confounders for the analysis. It is fundamental to take their effects into account, as not being aware of them could lead us to attribute their consequences to the variable we are interested in (here, gender).

In order to measure the effects of the genders of the judge and the applicant to the outcome, here defined as whether a judge voted in favor or against a finding of a violation. It is primordial to make sure the comparisons we make are reasonable. In simpler terms, we want to make sure we compare only similar units so that the outcome differences can be attributed to the genders of the judges (which we will refer to as treatment). Such "pairing," of the units is a

technique called “matching”. In short, the technique allows us to construct reasonable counterfactuals, pairs of male-female judges who are the most similar to each other (here, similarity is measured by how well the judges’ identity features, covariates, match). We then expect any differences in the conclusions they get to for similar cases as due to the judge’s gender identity.

After that, the author estimates the effects of judge gender using a Linear Regression model per each one of the issues, and for each issue one for male applicants, one for female applicants, and a last one for all applicants. Therefore, the model leads to 27 different regressions, from which 95% confidence intervals are estimated (for reference, a 95% confidence interval means we are 95% confident that the population parameter is between the given lower bound and upper bound). The results are then presented in Figure 1.

The presented coefficients represent the predicted difference in the probability of a vote in favor of finding a violation for a female versus a male judge. The main findings include:

1. Female judges are 8–9 percentage points more likely to favor the applicant than their male counterparts.
    a. Thus, there is some evidence that female judges are on average more liberal (in the sense that they favor the individual rights appeal more often), even after matching on and controlling for government ideology, national, and personal characteristics.
2. There is no evidence that female judges are overall more favorably disposed towards female applicants than are male judges. The coefficient on female

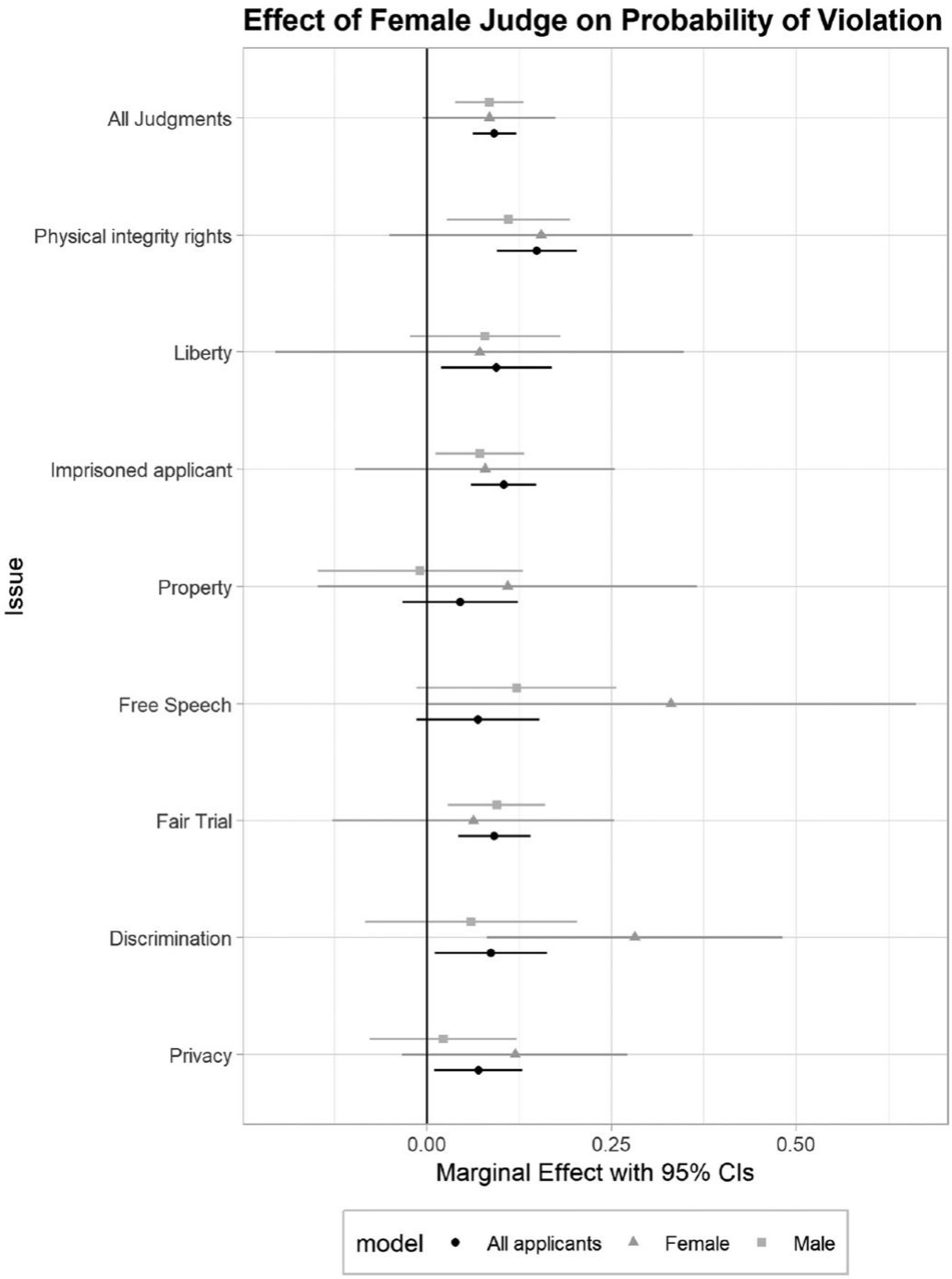

Figure 1. Regression coefficients post-matching, estimated effect of being a female judge on different subsets of judgments.

applicants for all judges is not significantly different from the coefficient on male applicants.

3. The only issue area where there is evidence for a gender applicant specific effect is discrimination issues (about 25 percentage points), it seems that female judges are more sympathetic than male judges to the types of discrimination cases filed by women.
4. Female judges are 15 percentage points more likely than male judges to find in favor of physical integrity rights. Moreover, female judges are 11 percentage points more likely to favor applications from detained individuals than their male counterparts (Voeten, 2020).

It is clear, then, that the judge and applicant's genders affect the decisions taken over the case. Nevertheless, it is important to point out that these results may not be generalizable since the sample size ends up being quite small and not representative. This because of the disproportional number of male judges in comparison to female judges. While matching, we match one male judge for each female judge, making the data set far smaller.

Action Plan

We propose and carry a partial replication of the study, in which we apply the same procedures to a new treatment. This time we will look for the effects of the judge identifying as religious or not to the probability of a vote in favor of finding a violation. We will apply genetic matching to the data set just as Voeten does. However, this time we will change the settings of the genetic matching model to take 'religious' as treatment instead of 'female,' as the study describes. Then, we will also change the definiton of each one of the 27 linear regression models

for them to take the variable 'religious' in the place of the 'female,' just as before. Following the code of the study, it is straightforward to plot the results in a similar fashion (See Figure 2).

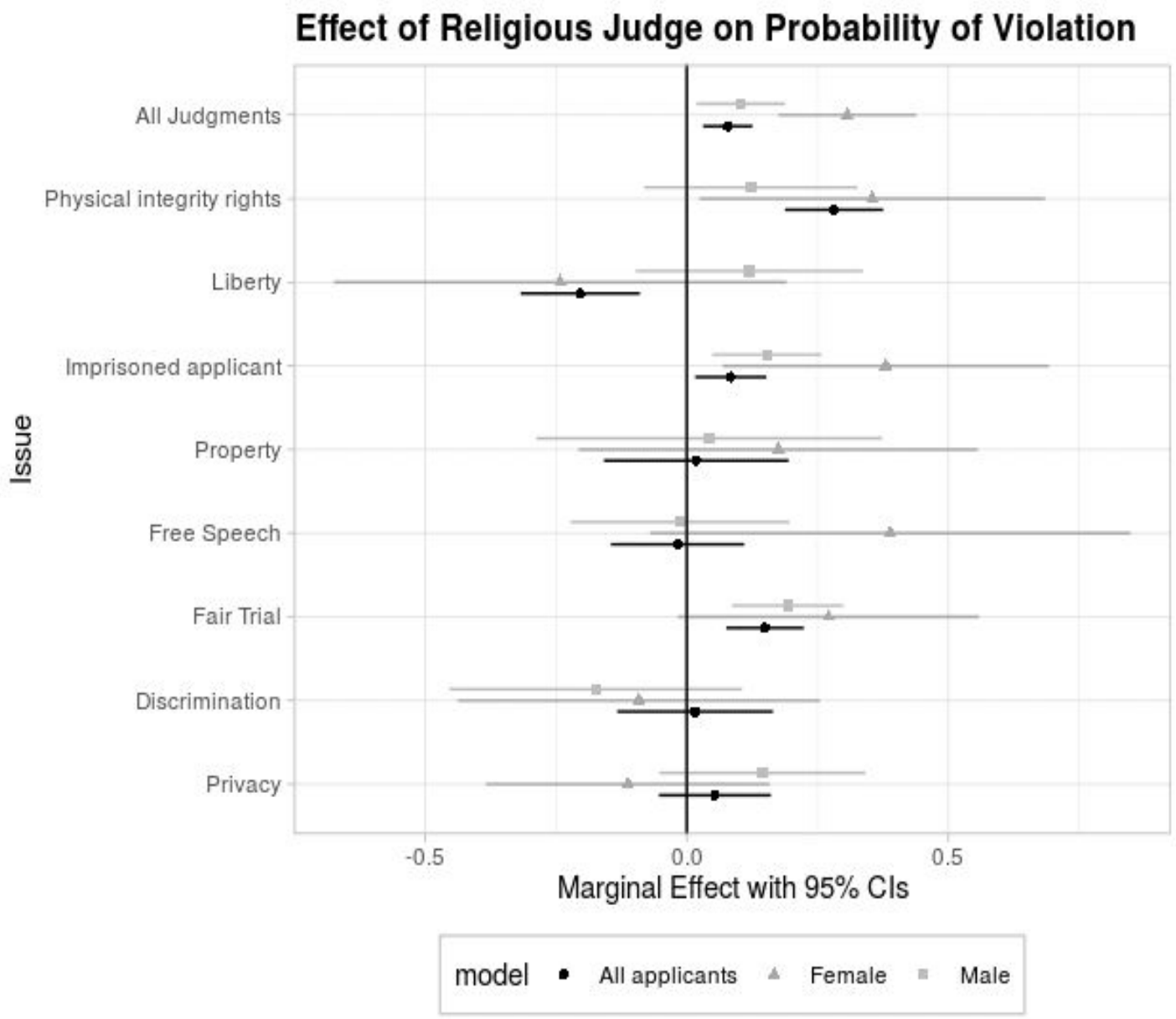

Figure 2. Regression coefficients post-matching for partial replication, estimated effect of being a religious judge on different subsets of judgments.

Note that for the case of this partial replication, the variable ‘female’ becomes part of the covariates, while ‘religious’ leaves the covariate set and becomes the treatment. This is a reasonable change since we can assume that the rest of the covariates (other identity traits) for gender will also be confounders for the effects of religious beliefs in court outcomes.

We can analyze these new results in a similar fashion to Voeten. As seen in Figure 2 the differences between the probabilities of a religious judge to favor a male or female applicant are greater than for the case of female vs male judges, which means being a religious judge vs not religious affects more the outcomes than being male vs female:

1. There is a difference of more than 20 percentage points for the religious judge to be in favor of a female applicant than a male across all judgements (top of the graph). In other words, being a female makes it 20 percentage points more likely for the religious judge to be on your favor than if you were a male.
2. The last trend can be found across many different issues (physical integrity rights, imprisoned applicant, property, free speech, fair trial and discrimination. This means that, generally, being a female increases the probability of the religious judge being in your favor.
3. Conversely, regarding privacy and liberty rights, being a woman actually decreases the probability of the judge being in your favor while applying for court..

The results of the partial replication, just as the results for the original paper support the idea that identity traits such as gender and religion will bias the decision process of a judge while being in favor of female vs male applicants.

In addition, it is clear that judges with different identities will have different biases, which brings up the question of whether or not it is possible to have a completely fair and objective judge. Similarly, it is interesting to remark that an effective solution for court subjectivity does not rely on increased diversity. Changing or 'fixing' the demographic will not solve the problem but only introduce different biases, hinting at the need for deeper, structural changes of the system.

On a different vein, and to conclude the analysis of the replication. It is crucial to point out a main concern with the reliability of the results of the partial replication. In the original study, the team ran a the genetic matching algorithm over 10 maximum generations of 1000 population size each. During the study replication, it was not possible to run the function with those parameters due to time and computer power constraints (it would take over two hours to run and the kernel would crash before it ended). Therefore, the study replication used a maximum of 10 generations but with a population size of 100, which still led to good balance of covariates and allowed us to produce the presented results.

Conclusion

Judicial proceedings tend to be unfair and biased. The evidence presented in the original paper and the replication support the idea that identity traits of judges affect the decisions they take depending on their preconceived notions of the applicants (for instance females vs males). We are able to use statistical methods like genetic matching and regression to measure such treatment effects. It is clear then, that deep, structural changes of the system are required, diversifying the judges' demographic would not solve the problem but only introduce different

biases. It is fundamental to advocate for and support profound jurisdictional reforms that can make of courts safe, reliable institutions.

References

1. Voeten, E. (2020). Gender and judging: evidence from the European Court of human rights. Journal of European Public Policy. Retrieved from: https://doi.org/10.1080/13501763.2020.1786146
2. The code used for the partial replication can be found here: https://gist.github.com/RafaelGA8/bfecfaf3aea0e7e1de48f2f8c0cc8c5f