ARTICLE

THE EFFECTIVENESS OF BAYESIAN JURY INSTRUCTIONS IN MITIGATING THE DEFENSE ATTORNEY'S FALLACY

Michael Conklin[*]

ABSTRACT

A previous study found that the effects of the prosecutor's fallacy can be minimized if juries are presented with a Bayesian jury instruction. This study examines whether a similar instruction can likewise combat the more pervasive defense attorney's fallacy or if the results of the previous study were more a function of increased confusion rather than an increased understanding of conditional probabilities. Additionally, prior performance in a statistics class is considered as a measure of susceptibility to this probabilistic fallacy.

TABLE OF CONTENTS

I. INTRODUCTION 74

II. HYPOTHESIS 75

III. STUDY 76

IV. RESULTS 77

V. POTENTIAL CRITICISM 78

VI. FUTURE RESEARCH 79

VII. CONCLUSION 80

[*] Assistant Professor of Business Law at Angelo State University.

APPENDIX A 81

I. INTRODUCTION

The defense attorney's fallacy is a lesser-known variation of the more common prosecutor's fallacy. The prosecutor's fallacy occurs when someone incorrectly believes that the probability of a random scientific match is equal to the probability that the defendant is innocent. To illustrate this, imagine a case where the defendant's blood matched that of the perpetrator of a crime. This blood type is so rare that there is only a 1 in 1,000 chance of someone possessing it. If you interpret this statistic to mean that there is only a 1 in 1,000 chance that the defendant is innocent, you are a victim of the prosecutor's fallacy. You improperly conflated random match probability (the frequency of the blood profile) with source probability (the probability that someone other than the defendant is the source of the blood).[1] In a town of 1,000,000 people, there would be roughly 1,000 people with the rare blood profile. Therefore, while it is true that there is only a 1 in 1,000 chance that a random person's blood would match the criminal's, the odds that a person who matches the rare blood profile is innocent—based solely on the matching evidence—is actually 999 in 1,000.

The defense attorney's fallacy is a variation of the prosecutor's fallacy where the jury is encouraged to disregard associative evidence as irrelevant, regardless of how rare the matching characteristic is.[2] To illustrate, imagine a trial with the same blood match as before but with additional evidence against the defendant, such as an eye witness and possession of the instrumentality of the crime. A defense attorney might nevertheless tell the jury that there is only a 1 in 1,000 chance of the defendant's guilt. While this would be correct if the prosecution's case was based solely on the blood match, the defense attorney is asking the jury, in effect, to disregard all evidence except the blood match. This is the defense attorney's fallacy. If the defense attorney is successful in getting the jury to believe this defective logic, the additional blood-match evidence can work to

1. William C. Thompson, *Letter to the Editor—The Prosecutor's Fallacy in George Clarke's* Justice and Science: Trials and Triumphs of DNA Evidence, 54 J. FORENSIC SCI. 504, 504 (2009), available at https://www.deepdyve.com/lp/wiley/letter-to-the-editor-the-prosecutor-s-fallacy-in-george-clarke-s-0aNb4JpdgN.

2. William C. Thompson & Edward L. Schumann, *Interpretation of Statistical Evidence in Criminal Trials: The Prosecutor's Fallacy and the Defense Attorney's Fallacy*, 11 L. & HUM. BEHAV. 167, 171 (1987), available at https://link.springer.com/content/pdf/10.1007%2FBF01044641.pdf.

erroneously reduce the jury's perception of guilt.

While the effects of the prosecutor's fallacy may be more severe than that of the defense attorney's fallacy (wrongful conviction versus wrongful acquittal), juries may be more susceptible to the latter than the former. One study found that the defense attorney's fallacy was more than twice as likely to be erroneously believed than the prosecutor's fallacy.[3] Additionally, when provided with wording to illicit the prosecutor's fallacy and the defense attorney's fallacy, research subjects were far more susceptible to the defense attorney's fallacy.[4]

A previous study (hereinafter "Rowe study") found that the effects of the prosecutor's fallacy can be minimized if juries are presented with a Bayesian jury instruction.[5] This present study examines whether a similar Bayesian jury instruction can likewise combat the defense attorney's fallacy. Additionally, prior performance in a statistics class is considered as a measure of susceptibility to this probabilistic fallacy.

II. HYPOTHESIS

The hypothesis going into this research is that the reason the Bayesian jury instruction was effective at combatting the prosecutor's fallacy in the Rowe study is not that the mock jurors actually gained knowledge about conditional probabilities after reading it. Rather, the 355-word instruction[6] was so complex that they simply became confused. This confusion about Bayesian probabilities was then mistaken for a lack of certainty as to the defendant's guilt, therefore resulting in decreased estimates of

3. *Id.* at 178 (68.5% of participants incorrectly labeled the defense attorney's fallacy as correct, while only 28.8% incorrectly labeled the prosecutor's fallacy as correct).

4. *Id.* at 178–79 (3% of participants' adjustments of probability of guilt was consistent with the prosecutor's fallacy while 56% was consistent with the defense attorney's fallacy).

5. Brenda Rowe, *A Possible Solution for the Problem of Juries Slighting Nonscientific Evidence: A Bayesian-Like Judicial Instruction*, 24 AM. J. CRIM. L. 541, 564 (1997). (Four different versions of a case summary and jury instruction combination were randomly distributed to participants. The case summary was either a weak version where the defendant was one of five hundred suspects or the strong version where the defendant was one of four possible suspects. The participants either received a Bayesian jury instruction after the case summary or no instruction. Therefore, participants received either 1) strong case no instruction, 2) weak case no instruction, 3) strong case with instruction, or 4) weak case with instruction. The findings demonstrated that, without the addition of the Bayesian instruction, participants did not properly take into consideration the base rate probability, thus succumbing to the prosecutor's fallacy (the non-instruction groups were more likely to think the defendant who was one of five hundred possible suspects was guilty than the defendant who was one of four possible suspects [everything else in the case summaries was constant]). Conversely, the groups that received the Bayesian instruction did, more properly take into consideration the base rate probability, thus demonstrating a lack of susceptibility to the prosecutor's fallacy.)

6. Rowe, *supra* note 5, at 551–52.

guilt, which happen to coincide with not succumbing to the prosecutor's fallacy.

If this hypothesis is correct, a Bayesian jury instruction should have the opposite effect than in the Rowe study when used in an attempt to combat the defense attorney's fallacy. This inverse result would occur because the prosecutor's fallacy erroneously increases perceptions of guilt (and therefore is combated by a reduction in perceived guilt). The defense attorney's fallacy, however, erroneously decreases perceptions of guilt (and therefore is combated by an increase in perceived guilt). Therefore, if the Bayesian jury instruction in this research causes a decrease in perceptions of guilt (thereby making the effects of the defense attorney's fallacy worse), then the hypothesis that the instruction does little to inform jurors about the proper application of conditional probabilities will be strengthened.

III. STUDY

The survey was administered to 103 undergraduate and graduate students at a regional university in the United States. After excluding two unusable submissions,[7] the remaining 101 were analyzed. In order to avoid asking a citizenship question, this study was not limited to jury-eligible participants. However, given the demographics of the group and that all were over the age of eighteen, the vast majority were likely jury eligible.

Participants in the study were randomly given one of two different versions of a survey. First, each version contained the same short case summary followed by a question asking participants to estimate the probability of guilt:

> A liquor store was robbed by a man wearing a ski mask. The store clerk was able to describe the robber's height, weight, and clothing, but could not see his face or hair. The police apprehended a suspect near the liquor store who matched the clerk's description but the suspect did not have the ski mask or the stolen money. In a trash can near where the suspect was apprehended, however, the police found the mask and the money.[8]

The second question involved the following additional information regarding an expert witness's testimony:

> Regarding the above robbery, a forensic expert testifies that samples of the suspect's hair were microscopically

7. One was excluded for being illegible, and the other was excluded for giving a range instead of an exact answer.

8. This is a slightly modified version of the language used to measure the defense attorney's fallacy in Thompson & Schumann's article. *See* Thompson & Schumann, *supra* note 2, at 173.

> indistinguishable from a hair found inside the ski mask. The expert states that only 2% of people have hair that would be indistinguishable from that of the defendant and stated that in a city of 1,000,000 people (the population where the robbery occurred) there would be approximately 20,000 such individuals.[9]

This wording was intentionally designed to be conducive to the defense attorney's fallacy. It attempted to downplay the non-scientific evidence by emphasizing how the defendant is just one of 20,000 people whose hair would also match that of the suspect.

The only difference in the two survey versions is that the first version contained only the case summary and additional hair-match evidence, asking the participant to make a percentage estimate of guilt after the first piece of information and then again after the second. The second version of the survey included a Bayesian jury instruction that accompanied the hair-match information in the second question. The instruction was a slightly abbreviated version of the Bayesian instruction that resulted in a reduction in the prosecutor's fallacy from the Rowe study (see Appendix A).[10] This instruction is based on the Bayesian mathematical model, which is the standard for accurately updating a prior probabilistic estimate when confronted with additional information.[11]

IV. RESULTS

The wording of the hair-match expert testimony—which was designed to illicit the defense attorney's fallacy—was effective. In the first version (no Bayesian jury instruction) 37.3% of survey participants gave responses indicating they were victims of the fallacy. This was defined as when a participant did not increase their initial estimate of guilt after reading about the additional hair-match evidence.

Contrary to the hypothesis of this study, the Bayesian jury instruction did help combat the defense attorney's fallacy. Of the group that was administered version two of the survey (with Bayesian instruction), only 22% gave responses indicating they were victims of the fallacy.

One of the background questions the survey asked was what grade the participant earned in their statistics class (which, for

---

9. This is a slightly modified version of the language used to measure the defense attorney's fallacy in Thompson & Schumann's article. *Id.* at 173.

10. Rowe, *supra* note 5, at 551–52.

11. Jonathan J. Koehler & Daniel N. Shaviro, *Veridical Verdicts: Increasing Verdict Accuracy Through the Use of Overtly Probabilistic Evidence and Methods*, 75 CORNELL L. REV. 247, 255 (1990).

this group, was most likely STAT 241, Introduction to Business Analysis). Surprisingly, receiving an "A" in the statistics class corresponded with an increased likelihood of succumbing to the defense attorney's fallacy in both versions of the survey.[12]

This unexpected result should give pause to an attorney going into voir dire with the strategy of selecting jurors of a certain educational background as a proxy for their understanding of probabilistic issues. Perhaps this counterintuitive result is a function of overconfidence on the part of people who did well in a statistics class and, conversely, humility (and therefore skepticism) on the part of those who did poorly. However, it is important to note that twenty-six of the participants did not provide an answer to the statistics grade question (likely because they forgot what their grade was or had not yet completed the statistics course). Therefore, the results of the statistics grade correlation part of this survey suffer from a small sample size.

V. POTENTIAL CRITICISM

A potential criticism of this study is that the case summary is overly simplistic; real-life cases contain far more information for jurors to consider. While this is certainly true, there is no reason to believe that adding ancillary information—and therefore complicating the case—would somehow result in an increased ability of participants to recognize and properly adjust to the defense attorney's fallacy. The simplification provided in this survey isolates the variable being measured. Therefore, it is a strength, not a weakness. The increased complexity of a real criminal trial would likely only make the statistical issue less apparent, and therefore jurors would more likely fall for the fallacy.

Relatedly, one might attempt to criticize this study by pointing out that real cases are resolved by the consensus of deliberating juries, not by averaging individual responses done in private. Again, this criticism points to a legitimate difference, but it is unclear how it is relevant to this study's findings. All of the related studies analyzed regarding jury instructions to combat statistical misunderstandings used individual results instead of mock juries. Furthermore, conducting enough mock jury deliberations to accurately measure the effects in this survey would be prohibitive; a sample size of just twenty-five juries would

12. Of participants who reported "A's" in the statistics class, 42.9% fell for the fallacy in the non-Bayesian instruction group compared to 37.3% of participants who did not report an "A". In the group that received the Bayesian jury instruction, 33.3% of participants who reported "A's" in the statistics class fell for the fallacy compared to only 22% of those who did not report an "A".

require 300 participants.[13]

One final criticism that could be made against this research is that the participants' answers are dependent upon other probabilities than just those immediately apparent based on the non-scientific and scientific evidence. To name a few: the odds that the expert who testified about the hair match is mistaken or intentionally lying, the odds that the ski mask found was not the one used in the crime, the odds that the ski mask was the one worn by the robber but the hair found inside was not the robber's, and the odds that the defendant is the victim of an elaborate framing effort by the police. Since each of these has a non-zero probability of occurring, a rational juror would need to take them into consideration when making a probabilistic estimation of guilt. However, the only way this would have an effect on the findings of this study is if the product of those probabilities is greater than the product of their negation, which is almost certainly not the case here. Simply put, it is extremely unlikely that the additional evidence of a hair-match identification would logically lead to a decrease in the likelihood of guilt.

VI. FUTURE RESEARCH

A qualitative study would need to be conducted to determine if exposure to a Bayesian jury instruction causes a better understanding of conditional probabilities or if the promising results demonstrated in this study were more a function of some alternative explanation.

Future research could include a similar study with judges as participants. Judges, even appellate justices, have been the victim of the prosecutor's version of the fallacy.[14] Furthermore, studies have shown that judges are no better than jurors when it comes to succumbing to fallacies.[15]

The Bayesian instruction tested in this research resulted in a reduction in the defense attorney's fallacy from 39.2% to 22%. However, 22% is still a significant amount for such a powerful fallacy. Future research should be conducted regarding what additional measures could lead to a further decrease. Unfortunately, the simple solution of allowing the adversarial

13. This is to say nothing of the logistical issues of getting groups of volunteers to show up at the same time and having contingency plans for no-shows.

14. Thompson & Schumann, *supra* note 2, at 171 ("Several appellate justices also appear to be victims of this fallacy See, e.g., *People v. Robinson*, 1970.").

15. Dawn McQuiston-Surrett & Michael J. Saks, *The Testimony of Forensic Identification Science: What Expert Witnesses Say and What Factfinders Hear*, 33 L. & HUM. BEHAV. 436, 440 (2009). ("Most studies that have compared the fact-finding and decision-making of judges to that of jurors have found no differences . . . including[ing] the ability . . . to draw inferences from probabilistic evidence.") (alteration in original).

nature of a trial to address the problem through vigorous cross examination of experts (and presentation of opposing experts) is unlikely to be effective.[16]

Due to the small sample size of those answering the question about their statistics class grade and the self-reporting nature of the question, future research should be conducted into the connection between background factors such as academic performance and susceptibility to legal fallacies.

VII. CONCLUSION

This study, combined with the Rowe study, provides strong evidence that Bayesian jury instructions help reduce the effects of the prosecutor's and defense attorney's fallacies. It also found that related academic performance is unlikely to be a proxy for the ability to recognize the defense attorney's fallacy. Finally, the Bayesian instruction only reduces the effects of the fallacy. Further research should be conducted into additional measures that could be taken to eliminate it.

16. *Id.* at 439 ("Generally, these studies have found little or no ability of cross-examination to undo the effects of an expert's testimony on direct-examination, even if the direct testimony is fraught with weaknesses and the cross is well designed to expose those weaknesses.").

APPENDIX A

Bayesian jury instruction utilized in version two of this study

This trial involves scientific evidence. In reaching your conclusion about whether the defendant committed the crime, first consider the chance of guilt based on all available nonscientific evidence. Next, update that belief by considering the scientific evidence. If there is a match, the strength of that match is given by the incidence rate of the matching characteristic. By using this process you will arrive at a final determination (based on all the evidence) of the chance that the defendant is guilty. This is your "posterior belief" about the likelihood that the defendant committed the crime. This will guide you in reaching a verdict. It is important that you assess the evidence in this manner because the strength or weakness of your prior belief plays an important role in arriving at a posterior belief which gives sufficient weight to all the evidence. Do not consider the scientific evidence to the exclusion of the nonscientific evidence. The difference in the nonscientific evidence between two hypothetical trials (A and B) could lead to a different just verdict in trial A than in trial B even when the scientific evidence in both trials is the same; therefore, it is important that you keep in mind your prior belief that the defendant committed the crime (based on all the nonscientific evidence) when you are presented with scientific evidence and that you update that prior belief based on the scientific evidence in order to arrive at a posterior belief of the likelihood that the defendant committed the crime (based on all the evidence, both scientific and nonscientific). That posterior belief may help you to reach a verdict.