APPLYING BAYES REASONING ON DISCRIMINATION DETECTION

Shuxuan Luo*

Abstract: This research highlights the crucial role of discrimination detection in effectively implementing anti-discrimination laws. Our main goal is to advocate for the use of Bayesian reasoning as a powerful method for detection, providing a clear mathematical definition for lawyers. We emphasize the benefits of employing Bayesian reasoning in a legal context, such as addressing the Prosecutor's Fallacy, considering evidence dependencies, and accounting for hidden assumptions like "common sense." To apply Bayesian principles, we carefully examine variables aligned with anti-discrimination laws like the Civil Rights Act of 1964 and the Americans with Disabilities Act. These laws encompass important areas such as race, gender, religion, and disability. Following the guidance of the Community Relations Service's resource guide, we collect variables related to protected characteristics, plausible discrimination scenarios, and indicators of reasonable accommodations. Our approach aims to present a well-rounded framework for discrimination detection rooted in Bayesian reasoning, meeting legal requirements and acknowledging the complex nature of discrimination in society.

Keywords: Bayesian Reasoning, Discrimination Detection, Civil Rights, Data Collection, Statistics for Lawyers

* Law School, The University of Sydney

Table of Contents

INTRODUCTION

To support the practical implementation of anti-discrimination laws, we propose a data-driven method for detecting potentially discriminatory circumstances. As Christine Mann Darden visualized the disparities between promoted male and female employees that won her a well-deserved promotion at NASA,[1] we can see the power of data as a tool to obtain new information and to observe existing trends that some of us are already experiencing while others are not yet aware of. In this paper, we advocate not only that we should cumulatively collect more informing data as guided and specified by relevant anti-discrimination laws, but also that we should analyze such data through the Bayesian way of thinking.

In a nutshell, Bayesian reasoning uses both past experiences and newly observed evidence to update our prior beliefs into posterior beliefs, and this updating process can happen repeatedly in light of continuously collected data, with the previously updated beliefs becoming the new priors.[2] Even with uncertain and inaccurate starting priors, the Bayesian property of "swamping the priors" can push the posteriors to converge to the same value when the observations are huge enough.[3] The point is that Bayes provides us a mechanism for repeatedly updating our inaccurate priors so as to reach a more accurate posterior. As the Bayesian approach requires us to consider all relevant data and thus "not just those which may appear to support our preconceptions," a huge advantage of applying Bayes is to encourage us analysts to "have an open mind."[4] Under a legal context, this advantage translates into less biased legal decisions because the Bayesian reasoning avoids misinterpretations and miscalculations, both of which have led to grave consequences in past legal cases, and because it can expose hidden assumptions regarding our common sense, the appropriate treatment of witness testimonies, and dependencies among pieces of evidence.

In the following paper, we first establish the mathematical basis of the Bayesian analysis for a foundational understanding of the approach. We then apply the derived formulas to real legal examples and explain the advantages of using Bayes under a legal context. Finally, we give a head start on applying the Bayesian reasoning to anti-discrimination laws by listing informing variables extracted from the Civil Rights Act of 1964, the Immigration Reform and Control Act of 1986, and the Americans with Disabilities Act. Such variables, though incomprehensive, include the protection of race and color, ethnicity, national origin, gender, religion, and disability in various scenarios focused on by the Community Relations Service.

I. THE BAYESIAN METHOD FOR LEGAL PROFESSIONALS

Finding the cause of the observed results has been one of the most important topics in practicing law, whether it be convicting the right person in criminal trials or identifying the complicated situations in which laws such as antidiscrimination laws

---

[1] CATHERINE D'IGNAZIO & LAUREN F. KLEIN, DATA FEMINISM (2020).

[2] stochazesthai, *Answer to "What Exactly Does It Mean to and Why Must One Update Prior?,"* CROSS VALIDATED (2015), https://stats.stackexchange.com/a/166322 (last visited Sep 19, 2023).

[3] Allen B. Downey, *Think Bayes: Bayesian Statistics Made Simple*, 37 (2012), https://www.greenteapress.com/thinkbayes/thinkbayes.pdf (last visited Oct 9, 2023).

[4] Ajitesh Kumar, *Bayesian Thinking & Real-Life Examples*, ANALYTICS YOGI (Feb. 2, 2023), https://vitalflux.com/bayesian-thinking-real-life-examples/ (last visited Sep 19, 2023).

and technology laws have been implicitly breached. Such causal ties that judges, juries, and many other legal professionals are trying to determine involve whether the committer caused the evidence presented in trials and if any discriminatory behaviors caused the unjust situation of plaintiff. However, causal inference has not been an easy topic as we need to examine the strength of the proposed causal relationship, how rare the cause happens, and the many other plausible explanations to the observed effect. The Bayesian method provides a useful framework for legal professionals to perform causal inference because it accounts for all the above factors in a logical and coherent way. The Bayesian method would help legal professionals to move towards a more corrected answer while at the same time acknowledging the existence of uncertainty.

Suppose a two-stage experiment has been conducted. Bayes' Theorem essentially answers the question that, if we observed the outcome of the second stage, what the probability of a specific outcome of the first stage is.[5] However, since the first stage experiment had already occurred, some may immediately dismiss this Bayesian reasoning on the ground that the outcome of the first stage is an historical fact that should not be understood as a probability. The outcome, they may say, either happened or did not happen, and therefore should not be assigned a probability of anything in between.[6] In response to this philosophical conundrum, a Bayesian statistician would specify her definition of the probability of the outcome of interest in the first stage as the "uncertainty" regarding the occurrence of such outcome[7] from her own perspective. This narrow definition emphasizing one's own perspective is the reason why the Bayesian approach is often considered as a "subjectivist view of probability."[8] As the Bayesian approach provides a framework to quantify one's uncertainty, it's not surprising that it has been applied in legal proceedings, where the second stage is the evidence found and the first stage being whether the suspect committed the crime or whether any discrimination happened. While the outcome of whether the crime was committed by the suspect is a historical, fixed fact, we can see that trials and other legal measures are still being conducted to infer such outcome. Having shown the usefulness of Bayesian reasoning, we will derive Bayes' Theorem and its derivative forms below.

A. Bayes' Theorem

The Bayes' Theorem is itself a form of causal analysis, where it moves from being presented the evidence to calculating the plausibility of the proposed cause. We illustrate this foundational process with step-by-step details for legal professionals below.

Denote all possible outcomes of the first stage as A and A^c and all possible outcomes of the second stage as B and B^c. Suppose our outcome of interest is A, and we are given the result of the second stage to be B. By the "general conjunction rule," where we assume the dependencies between first and second stage and therefore use

[5] FRANCISCO J. SAMANIEGO, A COMPARISON OF THE BAYESIAN AND FREQUENTIST APPROACHES TO ESTIMATION 33 (2010), https://link.springer.com/10.1007/978-1-4419-5941-6 (last visited Aug 11, 2023).
[6] *Id.* at 34.
[7] *Id.*
[8] *Id.* at 35.

one to infer another, we obtain $P(A \cap B) = P(A) * P(B|A)$.[9] By the multiplication rule of conjunctional probabilities, $P(B) * P(A|B) = P(B \cap A) = P(A \cap B) = P(A) * P(B|A)$, from which we get the foundational form of Bayes' Theorem:

Equation 1

$$P(A|B) = \frac{P(A) * P(B|A)}{P(B)}$$

Applying this theorem to a crude process of legal decision-making, we write $\mathrm{P}(Guilty|Evidence) = \frac{P(Guilty)*P(Evidence|Guilty)}{P(Evidence)}$. We define $P(Guilty|Evidence)$ as *posterior probability*, posterior to the presentation of evidence, and $P(Guilty)$ as *prior probability*, prior to evidence.

B. Generalized Bayes: Multiple Possible Outcomes

We acknowledge the difficulty in directly applying the Bayes formula to real legal cases. We therefore divide up the formula into a generalized form with more individual pieces with the intention of providing a more specified thus more practical framework of applying the Bayes Theorem.

Building upon Equation 1, we now specify how to calculate $P(B)$ in a general form. Denote all possible outcomes of the first stage as $A_i,\ 1 \leq i \leq N$, and all possible outcomes of the second stage as $B_j,\ 1 \leq j \leq M$. Suppose our outcome of interest is A_i, and we are given the result of the second stage to be B_j.

Plugging into Equation 1, we first get $P(A_i|B_j) = \frac{P(A_i)*P(B_j|A_i)}{P(B_j)}$. As each outcome in the first stage may be associated with the outcome of B_j, we have the total probability of B_j to be $\sum_{k=1}^{N} P(A_k \cap B_j)$. By the "general conjunction rule,"[10] we get $P(B_j) = \sum_{k=1}^{N} P(A_k)P(B_j|A_k)$, a weighted average of conditional probabilities. We therefore obtain the generalized Bayes' Theorem:

Equation 2

$$P(A_i|B_j) = \frac{P(A_i) * P(B_j|A_i)}{\sum_{k=1}^{N} P(A_k)P(B_j|A_k)}$$

In Equation 2, considering the cases of A_k, where k does not equal to i, prompts the legal professionals to consider all plausible causes in terms of both their causal tie with the observed effect and the rarity of the causes themselves.

C. Likelihood Ratio

[9] Northern Kentucky University, *Probability*, https://www.nku.edu/~garns/165/ppt9_3.html (last visi ted Aug 12, 2023).
[10] *Id.*

Beside directly considering the probability of the evidence being caused by the proposed reason, a formula involving the likelihood ratio gives the legal professionals an alternative method that evaluates the hypothesis of a defendant being guilty against the standard case in which a defendant is assumed not guilty. We illustrate this calculation of the "odds" below.

We define the likelihood ratio (LR) as $\frac{P(Evidence|Guilty)}{P(Evidence|NotGuilty)}$. We now prove that $posterior\ odds = prior\ odds * LR$:

Given: $posterior\ odds = \frac{P(G|E)}{P(NG|E)}$, $prior\ odds = \frac{P(G)}{P(NG)}$, $LR = \frac{P(E|G)}{P(E|NG)}$

Proof: By Equation 1, $\frac{P(E|G)}{P(E|NG)} = \frac{P(E)*P(G|E)}{P(G)} * \frac{P(NG)}{P(E)*P(NG|E)} = \frac{posterior\ odds}{prior\ odds}$.

Therefore, we get

Equation 3

$$\frac{P(G|E)}{P(NG|E)} = \frac{P(G)}{P(NG)} * \frac{P(E|G)}{P(E|NG)}$$

$P(E|G)$ is called "prosecution likelihood" while $P(E|NG)$ is called "defense likelihood."[11]

Mathematically, this ratio determines how much and in what direction the evidence updates the prior odds. If the ratio is bigger than one, then the evidence made the odds of guiltiness higher. If less than one, then lower. If equal to one, then the evidence made no meaningful updates on the uncertainty of a suspect's guiltiness. The LR, as characterized by Fenton et al., "is therefore an important and meaningful measure of the probative value of evidence."[12]

D. Bayes' Theorem with a Conjunction of Evidence

Finally, we expand our one-effect scenario into the situation with multiple effects. As the legal cases always involve multiple pieces of evidence, we believe this adaptation would be more relevant for legal professionals.

Define $P[G|(E_1 \cap E_2)]$ as the probability of a suspect being guilty given two pieces of evidence. By the general conjunction rule, $P[G|(E_1 \cap E_2)] = \frac{P(G \cap E_1 \cap E_2)}{P(E_1 \cap E_2)} = \frac{P((G \cap E_1) \cap E_2)}{P(E_2)*P(E_1|E_2)} = \frac{P(E_2)*P((G \cap E_1)|E_2)}{P(E_2)*P(E_1|E_2)} = \frac{P((G \cap E_1)|E_2)}{P(E_1|E_2)}$. We therefore obtain:

[11] Norman Fenton, Martin Neil & Daniel Berger, *Bayes and the Law*, 3 ANNU REV STAT APPL 51 (2016), https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4934658/ (last visited Aug 11, 2023).
[12] *Id.*

Equation 4

$$P[G|(E_1 \cap E_2)] = \frac{P\big((G \cap E_1)|E_2\big)}{P(E_1|E_2)}$$

Equation 4 shows a formula for evaluating two pieces of evidence. We are aware that this formula is still crude in the sense that a directly applicable formula would involve many more pieces of observed effects, and in turn would be too complicated to be specified by hand. However, as noted by Fenton et al., there is technology available—the "Bayesian network algorithms"—that is advanced enough to deal with multiple causes and multiple effects and thus to perform comprehensive Bayesian causal analysis.[13]

E. Bayesian Causal Inference

Bayesian reasoning is a powerful method that can greatly enhance the process of discrimination detection because of its ability to iteratively and logically infer unlawful discrimination, the cause, based on the increasingly available evidence, its effects.

Applying Bayes' Theorem in Equation 1, we see that inferring discrimination given the evidence requires two components—the prior, $P(Discrimination)$, and the standardized probability of the evidence assuming discrimination as the cause, $\frac{P(Evidence|Discrimination)}{P(Evidence)}$. The first component exposes our prior belief, often referred to as our implicit bias, and therefore can help us "recognize when that bias may be inappropriate or even misleading."[14] The prior belief about the general prevalence of discrimination in our society is the base rate to be corrected by the presentation of evidence. The second component, which can be expanded into $\frac{P(Evidence|Discrimination)}{P(Evidence|Discrimination)*P(Discrimination)+P(Evidence|OtherCause1)*P(OtherCause1)+\cdots}$ according to Equation 2, requires a thorough evaluation of both how strong the causal tie between discrimination and the evidence is as well as "the likelihood that the evidence might be explained by something other than [the] preferred hypothesis."[15] In other words, the numerator calculates the probability that a discriminatory behavior would consequently lead to the evidence, while the denominator of the standardized likelihood incorporates the true positives and the false positives. Considering the false positive cases is particularly useful because there can be many explanations other than discrimination that caused the existence of the evidence, and if this is the case, then using the evidence to conclude discrimination would not constitute a strong inference. The two components together form the Bayesian method of inferring the underlying cause of unlawful discrimination. In addition, the base rate can be iteratively corrected as new evidence appears by plugging in the posterior rate calculated in the previous iteration. In Chilson's words, the Bayesian reasoning's foundational logic is that "new evidence can always help refine a prediction… [as] we absorb [them] in rigorous and

[13] *Id.*
[14] Neil Chilson, *Bayesian Analysis as a Framework for (Legal) Thinking*, GEORGETOWN LAW TECHNOLOGY REVIEW (2016), https://georgetownlawtechreview.org/bayesian-analysis-as-a-framework-for-legal-thinking/GLTR-11-2016/ (last visited Oct 26, 2023).
[15] *Id.*

principled ways while recognizing that 100% certainty is rarely, if ever, warranted."[16]

II. THE ADVANTAGES OF APPLYING THE BAYESIAN REASONING UNDER A LEGAL CONTEXT

A. Avoiding Misinterpretations

The frequentist approach, a method of inference using a hypothesis test that involves concepts of the p-value and the confidence interval, is primarily used in legal proceedings.[17] However, interpreting p-values and confidence intervals correctly is quite challenging. In fact, a p-value, which calculates "the probability of observing the evidence given a hypothesis" $P(E|NG)$, "is often wrongly interpreted as being the same as the probability of the hypothesis given the evidence" $P(NG|E)$.[18] From Equation 1, we can quickly see why Bayes' Theorem can always properly avoid equating a pair of reversed conditional probabilities. The frequentist approach to hypothesis testing is also problematic because its confidence intervals "are almost invariably misinterpreted since their proper definition is both complex and counter-intuitive."[19] More specifically, a single 95% confidence interval cannot be interpreted as having a 95% chance of covering the true value of a parameter. Instead, the construction of 95% confidence intervals involves repeatedly calculating the intervals under a valid setting, and the resulting sequence of confidence intervals has the property that, on average, 95% of those intervals cover the true value.[20] In other words, any correct interpretations regarding probabilities of containing the true value in confidence intervals must involve "a long sequence of [such] intervals computed [repeatedly] from valid models," while we cannot quantify such probabilities in "any single [realized] confidence interval."[21] On the other hand, the Bayes version of an interval quantifying uncertainties, called a credible interval, has a much simpler interpretation where a single 95% interval calculated from one sample data has "a 95% probability that the true estimate would lie within [this] interval."[22] In seeking the probability of the true value being contained within a specific interval, we should thus calculate a Bayesian credible interval instead of a frequentist confidence interval.

In the context of legal proceedings, misinterpreting p-values to be $P(NG|E)$ when it should mean $P(E|NG)$, under the null hypothesis of not guilty, constitutes the "statistical reasoning error" of the Prosecutor's Fallacy.[23] In other words, "the chance of a rare event happening" does not equal to and cannot be used to infer "the chance of

---

[16] *Id.*
[17] Fenton, Neil, and Berger, *supra* note 12.
[18] *Id.*
[19] *Id.*
[20] Sander Greenland et al., *Statistical Tests, P Values, Confidence Intervals, and Power: A Guide to Misinterpretations*, 31 EUR J EPIDEMIOL 337 (2016), https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4877414/ (last visited Aug 18, 2023).
[21] *Id.*
[22] Luiz Hespanhol et al., *Understanding and Interpreting Confidence and Credible Intervals aroun d Effect Estimates*, 23 BRAZ J PHYS THER 290 (2019), https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6630113/ (last visited Aug 18, 2023).
[23] Bayes' Theorem in the Court – the Prosecutor's Fallacy : Networks Course blog for INFO 2040 /CS 2850/Econ 2040/SOC 2090, https://blogs.cornell.edu/info2040/2018/11/28/bayes-theorem-in-the-court-the-prosecutors-fallacy/ (last visited Aug 18, 2023).

a suspect's innocence."[24]

The infamous example where this fallacy was presented in a seemingly convincing argument that resulted in the wrong conviction of the defendants is People v. Collins.[25] In this case, a mathematician demonstrated an extremely low probability of possessing all characteristics described by the witnesses by applying the "product rule" to the estimated probabilities of each of the characteristic. Given the "one chance in 12 million" probability of observing a couple fulfilling all distinctive characteristics, the rarity of evidence was used to infer that "there could be but one chance in 12 million that defendants were innocent."[26] By using the low probability of observing the evidence, or a low "defense likelihood,"[27] to infer a low probability of innocence, it's clear here that the Prosecutor's Fallacy has been committed.

The Sally Clark Case is another infamous example demonstrating the Prosecutor's Fallacy. After Clark's first baby, Christopher, died, her second baby Harry died within two years. It was initially argued that "two unexplained infant deaths (SIDS deaths) are extremely rare in a family such as the Clarks."[28] Specifically, the probability was calculated to be "1 in 75 million."[29] However, as pointed out in the article by the Center for Statistics and Applications in Forensic Evidence (CSAFE), "saying that there is a 1 in 73 million chance that the babies died of SIDS is not the same as saying that there is a 1 in 73 million chance that the mother did not kill them."[30]

The Bayesian thinking can effectively prevent the occurrence of the Prosecutor's Fallacy, both by directly pointing out the difference between a set of reversed conditional probabilities shown in Equation 1, and by introducing the Likelihood Ratio shown in Equation 3 that enables the jury and jurors to evaluate the probative value of a piece of evidence based on both sides of stories. In Sally Clark's case, it was later calculated that "the probability that two infants [being] murdered in the same household is just 1 in 2 billion," reducing the odds of Clark being guilty down to only "4%," meaning that it's "much more likely (about 95% more likely) that the babies died of SIDS than that they were murdered by their mother."[31] As advocated by the CSAFE, "all parties must remember that there are always two sides to the story… [we] need to consider that probability of the evidence under the two competing hypotheses: the suspect is guilty or the suspect is not guilty."[32] What the Center presented in its calculation is exactly a Bayesian argument involving the Likelihood Ratio.

---

[24] scarraher, *Misuse of Statistics in the Courtroom: The Sally Clark Case*, CENTER FOR STATISTICS AND APPLICATIONS IN FORENSIC EVIDENCE (Feb. 16, 2018), https://forensicstats.org/blog/2018/02/16/misuse-statistics-courtroom-sally-clark-case/ (last visited Aug 22, 2023).
[25] People v. Collins, JUSTIA LAW, https://law.justia.com/cases/california/supreme-court/2d/68/319.html (last visited Aug 18, 2023).
[26] *Id.*
[27] Fenton, Neil, and Berger, *supra* note 12.
[28] scarraher, *supra* note 25.
[29] *Id.*
[30] *Id.*
[31] *Id.*
[32] *Id.*

B. Preventing Miscalculations

Miscalculations stemming from a failure to consider dependencies between different pieces of evidence is also a source to injustices in legal proceedings. The frequentists, by restricting its Product Rule into only calculating conjunctional probabilities of independent events,[33] acknowledge that the conjunctional probabilities of events whose occurrences influence each other are different from those of independent events. However, the frequentists' Product Rule has been inappropriately applied in courtrooms when treating multiple pieces of evidence, leading to miscalculations of probabilities that in turn resulted in unjust trial decisions.

Going back to the People v Collins example, another fatal statistical error pointed out in the appeal decision is "an inadequate proof of the statistical independence of the six factors" (namely, the inter-racial couple, beard, mustache, etc.).[34] The lack of a solid proof of mutual independence between pairs of factors makes the application of the Product Rule baseless and the results produced "erroneous and exaggerated."[35] The same type of miscalculation also occurred in the Sally Clark example, where the "1 in 75 million" estimate came from squaring the UK's crime statistics of "1 in 8500" "incidence of SIDS in the UK… in middle-class families with no known risk factors."[36] The square calculation came from applying the Product Rule, and we therefore can see the implicit assumption of the independence between the two deaths. However, there is no sound basis to assume the two deaths to be random events, as there have been several studies that "showed recurrence rates [of cot death to be] about five times the general rate,"[37] implying a higher probability of the second death given the first death.

The Bayesian approach, by calculating conditional probabilities, ensures the consideration of dependencies between events, and thus would by default avoid miscalculations due to wrong assumptions on independence. More specifically, by Equation 4, the term in the dominator, $P(E_1|E_2)$, would force the statistician to ponder upon the relationship between the two pieces of evidence.

C. Exposing Hidden Assumptions

A fixed rule of probabilities is that a conjunction of events always has a probability lower than or equal to that of each of the events, as illustrated in the famous "Linda the banker" example.[38] This rule gives rise to what Cohen called a "paradox" that leads to "inconsisten[cies]" between "probabilistic reasoning" and "legal reasoning."[39] More concretely, Cohen gave an example of two witnesses who independently provided their piece of evidence of witnessing events A and B, respectively. Assuming both witnesses to be pretty reliable, quantified with a

---

[33] Northern Kentucky University, *supra* note 10.
[34] People v. Collins, *supra* note 26.
[35] *Id.*
[36] scarraher, *supra* note 25.
[37] Stephen J Watkins, *Conviction by Mathematical Error?*, 320 BMJ 2 (2000), https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1117305/ (last visited Aug 22, 2023).
[38] Yong Lu, *The Conjunction and Disjunction Fallacies: Explanations of the Linda Problem by the Equate-to-Differentiate Model*, 50 INTEGR PSYCHOL BEHAV SCI 507 (2016), https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4967104/ (last visited Sep 1, 2023).
[39] Fenton, Neil, and Berger, *supra* note 12.

probability of $P(A) = P(B) = 0.7$, the resulting conjunctional probability of $P(A\&B)$ is reduced to only 0.49.[40] From the probability figures, we obtained a result that mathematically makes the suspect seem less guilty. However, from a legal perspective, the two independent, pretty reliable witnesses both testifying against the defendant should have positively supported the prosecutor's side. To reconcile the probabilistic and the legal reasonings, the use of Bayesian analysis can solve the above inconsistency by carefully accounting for more, if not all, of the hidden assumptions.

The problem with the above paradox lies in a direct translation of the witnesses' reliability into the probability of each event, even though such probability depends both on the reliability of the witnesses and on whether each event actually happened.[41] Consider the following example illustrated by Dawid. Witness 1 testified that "the sun rose today," while witness 2 testified that "the sun moved backwards through the sky today." By common sense, "we would, after hearing [those] testimony[ies], generally take the [probability of the first event to be] close to 1 and [that of the second event to be] close to 0," "irrespective of the general reliability of the witness."[42] In other words, even if witness 2 has a good record of telling the truth and thus has a reliability of 70%, one would still tend to believe that witness 2 made a wrong observation, meaning to assign the probability of the sun moving backwards to be close to zero. This example shows that the process through which we assign a probability to each piece of evidence is that we first have a baseline set by the common sense, and then we adjust the baseline probability based on witness testimonies. Using the Bayesian language, the baseline is a prior probability, and final probability of evidence is the posterior probability affected by both the prior and the reliability of the testimony.

The Bayesian approach accounts for the above three elements. Specifically, by applying Equation 3 onto individual pieces of evidence, we can write

Equation 5

$$\frac{P(A|a)}{P(A^C|a)} = \frac{P(A)}{P(A^C)} * \frac{P(a|A)}{P(a|A^C)},$$

where a is the presented testimony, A denotes that event A happened, and A^C denotes that event A did not happen. By characterizing the reliability of a witness in the Likelihood Ratio term, we reach a more intuitive interpretation of the reliability of a witness as "the probability that [the witness] will testify correctly, conditional on the true state of affairs."[43] Following Dawid's steps, we reconstruct Cohen's problem below.

Given that the witness is 70% reliable. Then $P(a|A) = 0.7$, and $P(a|A^C) = 1 - P(a|A) = 0.3$. Assign the prior $P(A)$ to have the probability of p, then $P(A^C) = 1 - p$. Plugging into Equation 5, we get $\frac{P(A|a)}{1-P(A|a)} = \frac{p}{1-p} * \frac{0.7}{0.3}$, which then can be simplified into

---

[40] *Id.*
[41] *Id.*
[42] A. P. Dawid, *The Difficulty About Conjunction*, 36 JOURNAL OF THE ROYAL STATISTICAL SOCIETY. SERIES D (THE STATISTICIAN) 91, 93 (1987), https://www.jstor.org/stable/2348501 (last visited Aug 28, 2023).
[43] *Id.*

$P(A|a) = \frac{7p}{3+4p}$. The posterior equals the reliability of the witness only if the prior p equals $\frac{1}{2}$. We now proceed to a conjunction of two pieces of evidence, a and b, for two events A and B. Suppose the two independent witnesses are each 70% reliable. That is, $P(a|A) = P(a^C|A^C) = P(b|B) = P(b^C|B^C) = 0.7$. Assuming events A and B are two independent events whose priors are both $\frac{1}{2}$, so as to recover Cohen's calculation of $P(A\&B|a\&b) = 0.49$. However, the evidence a and b actually increased the probability of guilt, from a starting point of the prior being $P(A\&B) = \frac{1}{2} * \frac{1}{2} = 0.25$.[44]

Applying Bayes' formula requires answers to questions of what our common sense is, what the precise relationship between the testimony and the event actually happening is, and whether or not the pieces of evidence are independent. All of those questions were not attended to and in turn implicitly assumed without a Bayesian analysis, thus driving to a seemingly inconsistent point between the fields of statistics and law. Bayes forces us to expose the above hidden assumptions, and therefore gives us more opportunities to face and in turn correct the biases in our reasonings. Recognizing the Bayesian reasoning's ability to incorporate implicit prior beliefs, we, together with many others, see the potential of applying Bayes in the area of antidiscrimination, where statistical evidence commonly plays a role.[45]

III. DATA COLLECTION

Having established a foundational understanding of the Bayesian analysis and of its ability to expose hidden assumptions and thus potential biases, we consider Bayes reasoning a great tool for detecting potential discrimination cases. We now proceed to concrete plans on data collection, which itself would be a step towards a bigger plan involving Bayesian data analysis.

Data to be used for updating prior beliefs through the Bayesian process shall be collected based on the framework proposed by the Community Relations Service (CRS), the "peacemaker" agency of the U.S. Justice Department, that itself was established under the Title X of the Civil Rights Act of 1964.[46] More specifically, we shall test for any discriminations against race, gender, religion, and disability[47] under the contexts of employment, work environment harassment, accommodations to employment, government services, public transit, businesses, and telecommunication, if applicable.

---

[44] . See also for a detailed calculation of the posterior probability when events A and B are not i ndependent, which is expected to be the usual case, by applying Equation 2. A special result achi eved is that the posterior has both a lower and upper bound, regardless of the prior. This result is consistent with the "swamping the priors" argument we presented in the introduction, in response to the criticism against the use of subjective priors in Bayes.

[45] Jason R Bent, *P-Values, Priors, and Procedure in Antidiscrimination Law*, 63 SSRN JOURNAL (2015), https://deliverypdf.ssrn.com/delivery.php?ID=932001119069127125016124029090113113046076048031004017071001066123027024102103001030124002121124043057052125120017103100010082106061094046072004084108120094067123095042050095006117104018007107124093006085104092123003080021102106077090111028006075110065&EXT=pdf&INDEX=TRUE (last visited Oct 29, 2023).

[46] Civil Rights Act (1964), NATIONAL ARCHIVES (2021), https://www.archives.gov/milestone-documents/civil-rights-act (last visited Sep 21, 2023).

[47] Michael L. Perlin & Valerie McClain, *"Where Souls Are Forgotten": Cultural Competencies, Forensic Evaluations, and International Human Rights.*, 15 PSYCHOLOGY, PUBLIC POLICY, AND LAW 257 (2009), http://doi.apa.org/getdoi.cfm?doi=10.1037/a0017233 (last visited Sep 14, 2023).

We identify the variables to be included cumulatively by probing through documents, as cited by CRS' resource guide, for each protected category below.

A. Race Discrimination

The race category, protected under the Civil Rights Act of 1964, includes race, color, national origin, and ethnicity.[48] By the U.S. Equal Employment Opportunity Commission's (EEOC) definition, race includes the race itself and the personal characteristics of "hair texture, skin color, [and] certain facial features," and color includes the "skin color complexion."[49] On the other hand, national origin and ethnicity include "[coming] from a particular country or part of the world," the "ethnicity or accent," or "appear[ing] to be of a certain ethnic background."[50] The "citizenship" and the "immigration status" aspects of national origin are also protected within the context of employment through the Immigration Reform and Control Act of 1986.[51] Finally, racial discrimination can occur due to an individual being "married to or associated with" a person of a certain race, color, national origin, or ethnicity.[52]

The scenarios in which discriminations against the race category are prohibited, as focused by the CRS resource guide, are "work situations," "harassment," and "employment policies," corresponding respectively to variables of "hiring, firing, pay, job assignments, promotions, layoff, training, fringe benefits, and any other term or condition of employment," "frequent or severe" remarks that result in a "hostile or offensive work environment," and "non-job-related [nor] necessary" employment policies that negatively impact a race category.[53]

E. Gender Discrimination

The gender category, also protected under the Civil Rights Act of 1964, has a wide variety of components that can be roughly summarized into "sex, gender, gender identity, and sexual orientation."[54] According to the CRS and the American Psychological Association, sex means a person's "biological status" that can be treated as a categorical variable of three levels—"male, female, or intersex." Gender refers to "the attitudes, feelings, and behaviors that a given culture associates with a person's biological sex," where "gender normative" means "behaviors that are compatible with cultural expectations" while "behaviors that are viewed as incompatible" are called "gender non-conformity." Gender identity means "a person's deeply-felt, inherent sense of being… [a] male, [a] female, or something else." A person whose gender identity does not match their sex is referred to as "transgender," while a matching person is called "cisgender." Finally, sexual orientation refers to "the sex of those to whom a person is sexually and romantically attracted," where "attraction to… one's own sex" is "gay or lesbian" while "attraction to… the other sex" is "heterosexual," and

---

[48] *Id.* at 4.
[49] Race/Color Discrimination, US EEOC, https://www.eeoc.gov/racecolor-discrimination (last visite d Sep 14, 2023).
[50] National Origin Discrimination, US EEOC, https://www.eeoc.gov/national-origin-discrimination (last visited Sep 14, 2023).
[51] *Id.*
[52] Race/Color Discrimination, *supra* note 50; National Origin Discrimination, *supra* note 51.
[53] Race/Color Discrimination, *supra* note 50; National Origin Discrimination, *supra* note 51.
[54] Perlin and McClain, *supra* note 48 at 4.

"attraction to members of both sexes" is "bisexual."[55] As the CRS resource guide does not include additional scenarios in which the gender category is protected, our scenario variables remain the same.

F. Religion Discrimination

As the Title VII of the Civil Rights Act of 1964 defines religion in a very broad way to include not only the "organized religion such as Christianity, Judaism, Islam, Hinduism, Sikhism, and Buddhism" but also the "sincerely held religious, ethical or moral beliefs" that are "new, uncommon," or even "that seem illogical or unreasonable to others,"[56] it's difficult to have a comprehensive list of variables to be measured under the religion category. In addition, the Supreme Court has decided that "it is not a court's role to determine the reasonableness of an individual's religious beliefs."[57] In an effort to be as inclusive as possible, we therefore treat "religion" as a binary variable without specifying the components under this category. We also include an association variable for the religion category as the Act also protects the person "[being] married to or associated with an individual of a particular religion."[58]

The scenarios focused by the CRS in which religion discrimination is prohibited are employment, workplace harassment, and accommodations to work. The employment variables, in addition to the ones listed under the race discrimination section, include segregation "such as assigning an employee to a non-customer contact position because of actual or feared customer preference," differential security requirements like demanding more extensive background checks for Muslims, and any forced religious activities as a condition of employment.[59] There's no additional variables for workplace harassment. Common workplace accommodations include "scheduling changes or leave for religious observances" and "dress and grooming practices."[60] The focus, however, is on the reasonableness of any requested accommodations, determined by the concept of "undue hardship," which is newly defined in the Supreme Court case *Groff v. DeJoy* as "more than a de minimis cost."[61] While the "undue hardship" is highly context dependent, some practical standards involve the "type of workplace," "nature of the employee's duties," "identifiable cost of the accommodation" with respect to the size and operating cost of the employer, "number of accommodated employees," workplace safety/ security requirement, "effect on workplace efficiency," and negative effects on coworkers or infringement of rights of other employees such as requiring others to "unwillingly do more than their share of potentially hazardous or burdensome work."[62] We integrate those indicating but non-decisive variables into our dataset in an effort to signal due responsibilities or

---

[55] Key Terms and Concepts in Understanding Gender Diversity and Sexual Orientation Among St udents: (527502015-001), (2015), http://doi.apa.org/get-pe-doi.cfm?doi=10.1037/e527502015-001 (la st visited Sep 14, 2023); Perlin and McClain, *supra* note 48.
[56] Section 12: Religious Discrimination, US EEOC (2021), https://www.eeoc.gov/laws/guidance/sec tion-12-religious-discrimination (last visited Sep 14, 2023).
[57] *Id.*
[58] Religious Discrimination, US EEOC, https://www.eeoc.gov/religious-discrimination (last visited Sep 14, 2023).
[59] *Id.*; Section 12, *supra* note 57.
[60] Religious Discrimination, *supra* note 59.
[61] Section 12, *supra* note 57.
[62] *Id.*

potentially undue hardships for employment accommodations.

D. Disability Discrimination

The disability category, protected under the Americans with Disabilities Act (ADA),[63] has three components: 1) the actual disability, defined as "a physical or mental impairment that substantially limits one or more major life activities,"[64] where major life activities include "caring for one's self, performing manual tasks, walking, seeing, hearing, speaking, breathing, learning, and working" [65] while being "substantially limited" is taken to mean that major life activities "are restricted in the manner, condition, or duration in which [such activities] are performed in comparison with most people,"[66] 2) a "history or record" of disability,[67] 3) is "perceived by others" as having disabilities.[68] The actual disabilities include but are not limited to "cancer, diabetes, post-traumatic stress disorder, HIV, autism, cerebral palsy, deafness or hearing loss, blindness or low vision, epilepsy, mobility disabilities such as those requiring the use of a wheelchair, walker, or cane, intellectual disabilities, major depressive disorder, [and] traumatic brain injury."[69] To be inclusive and flexible, we apply the strategy used under the religion category to again treat the "actual disability," "record of disability," and "regarded as having disability" variables to be binary instead of creating an incomprehensive list of specific types of disability.

Much more scenarios, in addition to workplace discriminations, are focused on by the CRS resource guide under the disability category. The major five areas are: employment, state and local government services, public transit, businesses that are open to the public, and telecommunications.[70]

Title I of the ADA protects both a person with disabilities and a person associated with an individual of disability under the context of employment, [71] following a similar three-part structure of aspects of employment, workplace harassment, and accommodations to work. The aspects of employment and harassment both involve the same variables listed before. Common accommodations include but are not limited to "making the workplace accessible for wheelchair users, providing a reader or interpreter for someone who is blind or hearing impaired, [and] making a schedule change."[72] The focus of disability accommodations is also on whether there's any "undue hardships," the standard of which, however, is defined to be "requiring significant difficulty or expense" [73] that is much higher than that of religious

---

[63] Perlin and McClain, *supra* note 48 at 5.
[64] Introduction to the Americans with Disabilities Act, ADA.GOV (2023), https://www.ada.gov/topics/intro-to-ada/ (last visited Sep 14, 2023).
[65] COMMONLY ASKED QUESTIONS ABOUT THE AMERICANS WITH DISABILITIES ACT AND LAW ENFORCEMENT, https://archive.ada.gov/q&a_law.htm (last visited Sep 14, 2023).
[66] *Id.*
[67] Introduction to the Americans with Disabilities Act, *supra* note 65.
[68] *Id.*
[69] *Id.*
[70] *Id.*
[71] Disability Discrimination and Employment Decisions, US EEOC, https://www.eeoc.gov/disability-discrimination-and-employment-decisions (last visited Sep 14, 2023).
[72] *Id.*
[73] 42 USC 12111: Definitions, https://uscode.house.gov/view.xhtml?req=(title:42%20section:12111%20edition:prelim) (last visited Sep 23, 2023).

accommodations.[74] The undue hardships for disability accommodations are also context specific, with the indicating but non-decisive variables of "the nature and cost of the accommodation needed," "the overall financial resources of the facility," "the overall size of the business" with respect to "the number, type, and location of its facilities" and "the number of its employees," and "the composition, structure, and functions of the workforce."[75] We treat those factors the same way as the indicating variables for undue hardships for religious accommodations by integrating them into our dataset for signaling purposes. One thing to note is that disability accommodations are only required for individuals with "actual disability" or a "record of" disability, but not for the ones who are "regarded as" having disabilities.[76]

The second scenario in which disability discriminations are protected against by the ADA is the state and local government services,[77] which include "all programs, services, or activities of public entities, from adoption services to zoning regulation."[78] In a similar vein, the major aspect that the public entities need to focus on for disability nondiscrimination is the "reasonable modifications" of policies and procedures for accommodation.[79] Such accommodations are guided under the principle of "equal [or further] treatment" that provides people with disabilities a "fair and equal opportunity to participate" in public services.[80] The "reasonableness," on the other hand, limits any modifications as to not "altering the nature of a program, service, or activity" and to not posing "objective, actual risk[s]" that impede the "safe operation" of the activity.[81] We use the "reasonableness of service" variable to determine the applicability of otherwise required accommodations.

Several specific accommodations required by the ADA, all limited by the "reasonableness" standard, are policies addressing 1) service animals, defined to be "a dog that has been… trained to… perform [related] tasks for an individual with a disability" and 2) wheelchairs and other power-driven mobility devices, and practices of how to 3) communicate with people who have disabilities. Public entities must allow "service animals" and "mobility devices" designed primarily for people with disabilities into all areas where the public is allowed to go. Factors of whether to allow other mobility devices include "type, size, weight, dimensions, and speed of the device," "volume of pedestrian traffic," "facility's design and operational characteristics," "legitimate safety standards," and "risk of serious harm to the environment or natural or cultural resources." We use the above variables to signal "appropriate mobility devices" that are also required to be allowed in public entities. Finally, the ADA requires public entities to "take the steps necessary to communicate effectively" with people who have disabilities, where the steps must be "practical" but can be "flexible" to "fit the circumstances" of "the nature, length, and complexity of the communication as well as the person's normal method(s) of communication." We use those factors to decide

---

[74] Section 12, *supra* note 57.
[75] 42 USC 12111: Definitions, *supra* note 74.
[76] 29 CFR § 1630.9 - Not making reasonable accommodation., LII / LEGAL INFORMATION INSTITUTE, https://www.law.cornell.edu/cfr/text/29/1630.9 (last visited Sep 23, 2023).
[77] Introduction to the Americans with Disabilities Act, *supra* note 65.
[78] ADA Update: A Primer for State and Local Governments, ADA.GOV (2023), https://www.ada.gov/resources/title-ii-primer/ (last visited Sep 14, 2023).
[79] *Id.*
[80] *Id.*
[81] *Id.*

the value for the binary variable "effective communication."[82]

The third scenario of disability nondiscrimination, also protected under the Title II of the ADA addressing public services, is "public transit." The available general requirement is that the public transit systems must provide "an equal opportunity [for people with disabilities] to benefit from their services."[83] The fourth scenario, protected under the Title III of the ADA, applies to businesses that are open to the public such as "businesses and nonprofits serving the public" and "privately operated transit."[84] The specific requirements also regard "reasonable modifications" to areas including but not limited to service animals, mobility devices, and effective communications,[85] adding no new variables to our cumulative list. The fifth scenario applying to "telecommunication companies" under Title IV requires that they "must provide services to allow callers with hearing and speech disabilities to communicate"[86] without further details. The last scenarios covered by laws other than the ADA are housing and air travel, protected respectively under the "Fair Housing Act" and the "Air Carriers Access Act."[87]

E. Variables

Based on the above thorough analysis of the types of protected characteristics and the scenarios in which they are protected from discrimination by the Civil Rights Act and the Americans with Disabilities Act, as focused on by the Community Relations Service, we compile below all variables to be collected for our dataset.

1. Protected Characteristics

- Race: race, hair texture, skin color, facial features, national origin, part-of-the-world origin, ethnicity, accent, perceived ethnicity, citizenship, immigration status, race association

- Gender: sex, gender, gender normative/ gender non-conformity, gender identity, transgender/ cisgender, sexual orientation, gay/ heterosexual/ bisexual

- Religion: religion (broad and inclusive), religion association

- Disability: actual disability, record of disability, perceived disability, disability association

2. Scenarios

- Workplace: hiring, firing, pay, job assignments, promotions, layoff,

[82] *Id.*
[83] Introduction to the Americans with Disabilities Act, *supra* note 65.
[84] *Id.*
[85] ADA Update: A Primer for Small Business, ADA.GOV (2023), https://www.ada.gov/resources/title-iii-primer/ (last visited Sep 14, 2023).
[86] Introduction to the Americans with Disabilities Act, *supra* note 65.
[87] *Id.*

training, fringe benefits, segregation, differential security requirements, forced religious activities, other terms or conditions of employment; work harassment; discriminatory employment policies; reasonableness of accommodations; accommodations provided

- Public and private services: reasonable modifications to policies and procedures, allowing service animals, appropriateness of mobility devices, allowing mobility devices, effective communication

3. Signals for (Un)Reasonableness

- Of general accommodations: type of workplace, nature of the duties, identifiable cost, number of accommodated employees, safety/security requirement, effect on efficiency, infringement of others' rights, overall financial resources, overall size of the business, composition of the workforce, structure of the workforce, functions of the workforce, nature of the public service

- Of mobility devices for people with disabilities: type, size, weight, dimensions, and speed of the device, volume of pedestrian traffic, facility's design and operational characteristics, legitimate safety standards, risk of serious harm to the environment or natural or cultural resources

- Of effectiveness of communication for people with disabilities: nature, length, and complexity of the communication, the person's normal communication method

CONCLUSION

With the final goal of detecting discriminations in mind, our paper mainly made two efforts—we proposed repeatedly applying the Bayesian method as our testing mechanism, and we summarized an incomprehensive list of informing variables, as guided and specified by relevant anti-discrimination laws covering race, gender, religion, and disability, whose values are to be cumulatively collected for Bayesian data analysis.

Looking forward, there is much more to be done towards our end, including 1) a close look into the technicalities for carrying out Bayesian data analysis, 2) refining the appropriate types (categorical, continuous, binary, etc.) of each of the variable to fit the purposes of the corresponding laws, 3) thinking through the exact relationships between signaling variables and the accommodation variables, not to mention 4) the practical details for data collection and the data cleaning process.