

Statistical Evidence, The Proof Paradox, and Individualized Evidence

By Isabel Wang

AUTHOR BIO

Isabel Wang is a student of Horace Greeley High School. She is passionate about topics in philosophy and mental health advocacy. Her academic interests lie at the intersection of philosophy, psychology and law. Her interest in philosophy is inspired by conversations she has had in her teenage years, which gave her the opportunity to discuss philosophical topics with a variety of people, from her parents to other peers to teachers. Her first-hand witnessing of the mental health struggles of teenagers and young adults motivated her to study psychology. She is interested in studying the mass incarceration of the mentally ill, youth mental health and the moral obligations of a society.

ABSTRACT

The proof paradox has been a puzzling debate among philosophers: is generic statistical evidence enough in court for a conviction? This paper discusses the proof paradox and attempts to explain it away. It hypothesizes that statistical evidence does a poor job at representing the individual and therefore should not be used. The paper is written through many hypothetical examples, and credible research, including historical court cases. The research indicated that generic statistical evidence is indeed not sufficient evidence and often poorly represents an individual, stating that individualized evidence should be used in its place.

INTRODUCTION

The proof paradox, which is heavily debated, discusses the question of whether statistical evidence can be solely used in order to decide the outcome of a case (Gardiner, 2018; Ross, 2020; Webster, 2021). This paper will attempt to explain away the proof paradox, beginning with a case where statistical evidence was presented as evidence.

Nearly a century ago, Betty Smith accused the bus company, Rapid Transit, of running her off the road (*Smith v. Rapid Transit*, 1945). Smith provided a proposition that there was a high probability that the defendant owned the bus that was on the road the night she was run off, because the defendant owned the majority of the buses on that route. Eventually Rapid Transit won the case, and the court's decision rested on the fact that there was no direct evidence that the bus on the road that night belonged to the defendant. That is, although there was high probability that Rapid Transit did own the bus that ran Smith off the road, this merely statistical evidence was deemed insufficient to prove Rapid Transit guilty. However, this was a civil case and the requirement for the burden of proof was a preponderance of evidence, meaning an event is more likely to be true than not. Smith provided a probability that met this requirement, yet she lost her case (*SD 5*, 2024).

This case is a prime example of how (*merely*) statistical proof, like the probability that the defendant owned the bus on the road, may not be sufficient evidence to convict someone. Importantly, it has been noticed that the use of statistical evidence in legal reasoning leads to the *proof paradox* (see Ross, 2024). In this paper, the following questions will be investigated: How can we explain away the proof paradox and what does the paradox imply

for the use of statistical proof in different kinds of cases?

The paper proceeds as follows. First, an explanation of why the proof paradox is 'paradoxical' will be provided. The key to the explanation is that when generic proof is used for statistical evidence, it does an abysmal job at representing each individual, since individuals may have their own factors that determine whether this generic proof applies to them or not. Then it will be argued that because generic, non-individualized statistical evidence is poor evidence, it is only useful in cases where the stakes are low.

STATISTICAL EVIDENCE AND THE PROOF PARADOX

What is statistical evidence? Basically, statistical evidence uses probability or correlation to indicate how likely someone is guilty; it is the use of evidence that is solely statistical. The use of statistical evidence in law, however, is a heavily debated topic (Fried, 2022; Ross, 2020). An example of how statistical evidence is used in law is evident in *Smith v. Rapid Transit*. In that case, Smith provided evidence that Rapid Transit was the only bus company licensed to operate on the route she was on the night of her accident. According to the decision, because private buses that were unaccounted for could possibly have been operating the bus that ran her off the road, the court decided this was not sufficient evidence. Though the likelihood of a private bus being the bus that ran her off the road is highly unlikely, it is still a possibility that causes this to be an example of the use of statistical evidence (*SD 5*, 2024).

The use of statistical evidence can lead to peculiar consequences, such as the proof paradox. Consider the following two examples (from Ross, 2024) in order to understand the

proof paradox and how statistical evidence is used within it.

1. *The prisoner*: There are one hundred prisoners exercising in the prison yard. Ninety-nine of them suddenly join in a planned attack on a prison guard; the hundredth prisoner plays no part. There is no evidence available to show who joined in and who did not (Ross 2024). In this case, the high probable chance of any of these hundred individuals being involved is ninety-nine percent, if the sample group in question is the one hundred prisoners in the yard. This ninety-nine percent chance is determined using statistical evidence. The question now is whether this statistical evidence is enough proof to punish all the individuals in the prison yard, knowing one among the punished prisoners will be innocent. This is where the paradox comes in: on the one hand, because for each of the prisoners, their probability of being involved in the attack is 0.99, each one of them should be punished. On the other hand, we also know that not all of the prisoners were involved, and therefore, it is *not* the case that each one of them should be punished.
2. *The blue bus hypothetical*: This is a realistic example adapted from the *Smith v. Rapid Transit* case mentioned above. It goes that a bus negligently causes injury to a pedestrian, but it is not known which company the bus belongs to. On the route where the accident occurred, the Blue Bus Company runs 75 percent of the buses. There is no further information available to settle which company the bus belongs to (Ross 2024). In the blue bus case, the statistical evidence presented is the high

chance that the blue bus owns the bus involved in the accident, solely because they own a high percentage of buses on that road, which is the sample group. The paradox would ask whether this is sufficient evidence to punish the blue bus company.

Let us look at the blue bus case with a different perspective. 75 percent seems like a high percentage, but if this hypothetical case occurred four times, then it is possible that the blue bus did *not* cause one of these accidents. In that case, it would only be right to punish them for three of these accidents. But which three should the blue bus company be punished for? Who's to say the original hypothetical was not the *one* that wasn't their fault? On one hand, the fact that the Blue Bus Company owned 75 percent of the buses where the accident occurred, they should be punished for the accident. On the other hand, there is also a known 25% chance that the bus involved in the accident did not belong to them, so it would be unjust to punish them without any certainty that they owned the bus in question. It is important to consider that fact before choosing whether or not to punish them.

Consider another case, this time a real case that had occurred in 1968.

People v. Collins: This is a case from 1968 where statistical evidence was used in order to convict an interracial couple for a robbery. In this case, a woman was robbed by someone who she observed as having medium-blond hair. An eyewitness, John Bass, gave a testimony in which he identified the robber as caucasian with dark blond hair in a ponytail. Additionally, he identified a man in the partly yellow automobile in which the blond robber had been near during her escape. He described that man as a Black man who had both a mustache and beard. The couple, the white

SCHOLARLY DEBUT

by Scholarly Review

Fall 2025

woman and Black man, were convicted for robbery based on the rarity of their combination of characteristics, with the claim that it was unlikely that they possessed the full set of similar traits as the robbers. This claim was based on mathematical probability and statistics, as presented below, notably not proven beyond a reasonable doubt. The left side presents the characteristic, while the right displays the probability of each individual characteristic occurring (Fairley & Mostellert, 1974).

Partly yellow automobile - 1 out of 10
Man with mustache - 1 out of 4
Girl with ponytail - 1 out of 10
Girl with blond hair - 1 out of 3
Black man with beard - 1 out of 10
Interracial couple in car - 1 out of 1000

Combining all of these characteristics and probabilities, the court found that there was a 1 out of 12 million likelihood of this combination of traits (Fairley & Mostellert, 1974).

Considering this case, there was less than one percent of a probability of those combinations of traits, and so on the one hand, the couple should be convicted based on the low likelihood of their rare characteristics. However, when considering each of these individual traits, the possibility is not as low. Furthermore, they should not be convicted solely because of their body features and the color of their car. The convicted could not be proven to *be* the robbers beyond a reasonable doubt using clear evidence, and were solely accused because they matched the characteristics of the robbers. Though it is highly likely that the accused were the robbers, based on probability, they could not prove this fact beyond a reasonable doubt without more concrete evidence. They should not be convicted based on only the individual pieces of statistical evidence, which is not extremely personalized

nor rare and which does not take into account DNA and other such traits.

It is important to note that this conviction was later overturned, claiming that in this case, statistical evidence was used mathematically incorrectly and the resulting probability was not accurate, so it was not enough to prove guilt. An accurate depiction of the statistic would show that there was a 41% likelihood of another couple with the same straight (*People v. Collins*, 1968). This percent would still be enough to prove this conviction if it were a civil case, however it was not. In a criminal case, evidence had to prove that beyond a reasonable doubt, the defendant committed the crime.

In the prisoner case, the blue bus hypothetical and *People v. Collins*, the stakes of the case are high. Stakes are high in cases when the punishment will be severe and impact the people being punished to a greater extent. Since the stakes are high in these cases, generic evidence cannot be considered as a form of evidence. High stake versus low stake cases will be discussed later.

INDIVIDUALIZED EVIDENCE

In order to explain away the proof paradox, a distinction about legal evidence must be introduced: individualized versus generic evidence. Let us first consider *individualized* or individual-specific legal evidence. This type of evidence suggests an *appropriate causal relation* between the evidence and the individual being indicted (cf. Thomson, 1986; Gardiner, 2018); this causal relationship enables us to trace back to the specific individual who has committed the crime through available evidence. Put differently, an appropriate causal relation means the evidence has a unique connection with a certain individual, and, very importantly, will not be influenced by the choice of reference

SCHOLARLY DEBUT

by Scholarly Review

Fall 2025

class (detail below). Examples of individualized evidence are footprints, fingerprints, or blood samples that can point towards a direct causal connection between the criminal conduct and the suspect. These pieces of individualized evidence will be able to help us reconstruct the *causal* process through which an individual has committed a crime. This type of evidence, therefore, is not merely statistical, even though it still falls short of certainty. The lack of full certainty in this form of evidence will be addressed later.

To illustrate, consider a hypothetical example in which 99 gang members were involved in killing Y. Suppose that 100 people were witnessed to be at the crime scene (with one of them being an innocent bystander). The police found 99 fingerprints on the weapons and traced back to the 99 gang members. This would be the use of individualized evidence. Note that as a matter of fact, fingerprints are not 100% reliable; they may only be 0.99 (or less) accurate in identifying the true suspects. The rate of falsely identifying a positive fingerprint is around 0.1% (Ulery et al., 2011). Still, fingerprints are a better form of evidence than merely generic statistical evidence, since they are individualized to each defendant. Since this is an example with high stakes, it would be unjustifiable to use generic statistical evidence because of high consequences.

GENERIC EVIDENCE

Other than individualized evidence, there is generic evidence which probabilistically or statistically represents a group without looking at each individual. In fact, such evidence often uses probabilities that are based on oversimplified probabilistic models. For example, if individual A has feature F and it is known that a crime was committed by someone with feature F, individual A will have a certain probability of being involved despite having no

individual link to the crime. This is because individual A is put into a reference group of people with feature F, even though this feature has nothing to do with the individual's interior traits or any form of individualized evidence. Often, people are put in reference groups that don't represent them as an individual, and are grouped with the wrong class of people. They are put in the wrong reference class based on a feature trait, or because they are in the wrong place at the wrong time (cf. Colyvan & Regan, 2007). By doing so, they get compared to a group of people with whom they do not have many similarities and with whom they cannot be statistically compared in an accurate and individualized manner. This was the type of statistical evidence used above in both the prisoner case and the blue bus case.

As mentioned in the section above, using generic evidence, each prisoner has a 99% chance of having attacked the guard simply by being present at the crime scene; but this probability relies solely on the reference class being used, meaning the statistic will change if the reference class changes. Similarly, the blue bus company has a 75% chance of owning the bus involved in the evidence because they owned 75% of the buses *on that road*, which is also the reference class in question.

Generic evidence does not take into account many other individual factors that may influence the likelihood of involvement in either the attack or accident. For example, prisoner A may have a violent history of crime, which increases their likelihood of involvement. On the other hand, prisoner B may be the best behaved criminal in the yard, and a week away from being let out of prison. This would severely decrease the likelihood of prisoner B's involvement, since they would compromise their release by being involved with an attack. Additionally, in the blue bus case, if it is common knowledge that the company is always

SCHOLARLY DEBUT

by Scholarly Review

Fall 2025

very careful and rarely causes accidents, it would decrease the chances that they were involved in the accident. Alternatively, if their drivers are known to be reckless, it may increase the chances. All of these examples are the use of individualized evidence that looks at what individuals in the sample group are like and their backstory. To sum up: to explain away the proof paradox, individualized evidence should be used in place of generic evidence when it is available. This is especially important when the situation has high stakes.

EXPLAINING THE PROOF PARADOX: INDIVIDUALIZED VS. GENERIC EVIDENCE

Although individualized evidence does not guarantee full certainty when used and may result in similar percentages as generic evidence, it is more accurate and personalized than generic evidence. Fingerprints are a form of individualized evidence mentioned previously, but it is known that fingerprints *do not* give 100% accuracy. Let us say that a certain individual had a 99% fingerprint match to the fingerprints that were thought to belong to the killer in a certain murder. Returning to the 99 gang members who killed Y, any individualized evidence they had in that case will be hypothetically removed. So now, all 100 people who were at the scene of the crime have a 99% chance of being involved in the murder, since there is nothing that can differentiate who is in the gang and which one individual isn't. In both murder cases, the individual or individuals have a 99% chance of being involved in the murder. The first case uses individualized evidence, and the second uses generic, but the percentage conclusion is the same.

Some may ask what is the difference between these two examples if the percentage is the same? What makes it fair to easily punish the first individual but to hesitate in punishing all

100 gang members? The biggest difference is justified belief in the relation between the individual and the crime. The individual in the first case has a 99% chance of being involved in a murder based on evidence that directly points to him as an individual, while the second case is general. In the second case, it is as if they are all being dehumanized and they are just part of a certain equation. Changing the sample group would change the equation. In the first case, changing the sample group, if there was one, would not change the percent match of the fingerprint. However, the certainty that one of the 100 individuals in the second case is definitely innocent. This suggests that the problem with statistical evidence is not that the evidence is 'statistical', but that the evidence is not individual-specific. Such kind of evidence may judge someone guilty simply based on some of their superficial characteristics—by race or where they were at a certain time—but this evidence is not really about the individual.

In a case where two or more people have a 99% match to the fingerprint of a certain perpetrator, individualized evidence would also lead to the same immoral result as generic evidence. In any case where more than one person has an almost complete match to the DNA gathered connecting to the criminal, it is a guarantee that holding them both accountable will result in an unjust conviction. That is the exact same problem generic evidence results in. So in cases like these, how can we distinguish generic and individualized evidence? Since they both result in the same, unfair conclusion, there is no way we can use individualized evidence in a more moral way than generic evidence. That is untrue. First, DNA evidence is a form of evidence that is accepted because of how it directly relates to the autonomy of a person. Though, it is still not completely fair, it relates more to the individual persons than generic evidence can. Generic evidence simply generalizes individuals and puts them into

categories. Individualized evidence, as unclear as it may be at times, is still a result of traits directly taken from individuals. Second, punishment is not a result of only DNA evidence. There are other factors that impact the decision, like eye witnesses. These factors together will produce a clearer and more accurate result. That is why a fingerprint match, though similar in probability as generic evidence, can be used.

USING STATISTICAL EVIDENCE WHEN INDIVIDUALIZED EVIDENCE IS UNAVAILABLE

My solution to the proof paradox was that individualized evidence should be used in place of generic evidence when it is available. So what if individualized evidence isn't available? Should we use generic evidence? Returning to the stake of the cases, we should avoid using (or solely relying on) non-individualized statistical evidence when the stake of wronging someone is high. This is because punishment will be greater in these cases, and can ruin an innocent person's future. This innocent individual was simply at the wrong place at the wrong time.

In the prisoner case, stakes are high because it is a criminal offense that could lead to severe punishments. If absolutely no individualized evidence was available in the prisoner case, which is unlikely, then they should lessen the punishment and punish everyone, or leave the case open. Even when the generic evidence can prove beyond a reasonable doubt that the defendant was the perpetrator, such as in *People v. Collins*, generic evidence still should not be the only evidence used to provide this proof. In the case of the prisoners, under no circumstances should the severe punishment that usually comes with this attack be given to all 100 prisoners. This would be very unfair and unjust towards the one who was

innocent. The innocent prisoner will be punished anyways, with this lessened punishment, because the guard who was attacked needs to receive justice. By not punishing any of the prisoners at all, the guard would not get justice and the prisoners may plan another attack, since the first went unpunished. The burden of proof for such a criminal case is not enough to define the specific punishment.

Certain minor violations of traffic laws are considered low stakes because punishment will be less severe. These minor offenses can be punished without harming somebody too much if they were innocent in actuality. An example of this is charging someone for running a stop sign. If nobody is injured, it is easy to charge the owner of the car that did not stop at the stop sign if traffic cameras in the area caught it, whether or not the owner was actually in the driver's seat. If somebody is severely injured, it would be different, because it would become a greater violation with more drastic punishments. However, a minor case such as *Smith v. Transit* should require clear and convincing evidence to convict rather than a preponderance of proof. This would be more evidence necessary than a preponderance of proof but less than beyond a reasonable doubt. Although this may make it more difficult for the plaintiff to prove the defendant is guilty, it is more ethical to allow this method. It is morally wrong to punish an individual who is innocent, despite the dangers and unfairness of allowing a guilty defendant to go free.

In general, if only generic evidence is available, then minor offenses should be punished, and high stake cases can be punished with a lessened form of punishment (or should remain open). Generic evidence cannot be the sole form of evidence to prove beyond a reasonable doubt that a defendant committed the crime in a criminal case. In doing so, a wrongful

conviction may occur, leading to harsh punishments for an innocent individual.

CONCLUSION

To summarize, the proof paradox proves to be paradoxical because when generic, non-individualized, statistical proof is used as evidence, it does not accurately represent the individual. Because of this, generic statistical evidence should not be used if individualized evidence is available. Individualized evidence doesn't remove the statistical aspect of the evidence, but that isn't the issue with generic evidence. The issue is that the statistics in generic evidence don't relate to the individual. If individualized evidence is unavailable, generic evidence can be used in low stake cases. Especially in high stake cases, using generic evidence would risk severely punishing an individual who is actually innocent, which would be extremely unjust, as evident in the case of *People v. Collins*.

REFERENCES

Colyvan, M., & Regan, H. M. (2007). Legal decisions and the reference class problem. *The International Journal of Evidence & Proof*, 11(4), 274–285. <https://doi.org/10.1350/ijep.2007.11.4.274>

Fairleyf, W., & Mostellert, F. (1974). A conversation about Collins. *University of Chicago Law Review*, 41(2), 243–244. Chicago Unbound. <https://chicagounbound.uchicago.edu/cgi/viewcontent.cgi?article=3802&context=uclrev>

Fried, A. (2022). Statistical evidence and the criminal verdict asymmetry. *Synthese*, 200(6). <https://doi.org/10.1007/s11229-022-03947-w>

Gardiner, G. (2018). Legal burdens of proof and statistical evidence. *The Routledge Handbook of*

Applied Epistemology, 179–195. <https://doi.org/10.4324/9781315679099-14>

People v. Collins, 68 Cal.2d 319, (March 11, 1968). <https://scocal.stanford.edu/opinion/people-v-collins-22583>

Ross, L. (2024). *The philosophy of legal proof*. <https://doi.org/10.1017/9781009127745>

Ross, L. D. (2020). Recent work on the proof paradox. *Philosophy Compass*, 15(6), 1. <https://doi.org/10.1111/phc3.12667>

SD 5. (2024). Harvard.edu. <https://cyber.harvard.edu/torts01/syllabus/readings/SD5.html>

Thomson, J. (1986). Liability and individualized evidence. *Law and Contemporary Problems*, 49(3), 199–219. <https://scholarship.law.duke.edu/lcp/vol49/iss3/12/>

Ulery, B. T., Hicklin, R. A., Buscaglia, J., & Roberts, M. A. (2011). Accuracy and reliability of forensic latent fingerprint decisions. *Proceedings of the National Academy of Sciences*, 108(19), 7733–7738. <https://doi.org/10.1073/pnas.1018707108>

Webster, A. K. (2021). Proof Paradoxes, Agency, and Stereotyping. *Philosophical Issues*, 31(1), 355–373. <https://doi.org/10.1111/phs.12191>