Problems With Using Statistics to Justify Institutional Policies

Justin Shin
Bard College, js4249@bard.edu

Problems With Using Statistics to Justify Institutional Policies

A Senior Project submitted to
The Division of Social Studies
of
Bard College

By
Justin Shin

Annandale-on-Hudson, New York

May, 2017

Abstract

It is becoming increasingly common for institutions to use statistics to inform policy decisions. We should be prepared to ask ourselves what regulatory principles should be imposed on institutions that seek to justify certain policies through deference to a statistical analysis. This paper will examine the difficulties that come with using statistics to justify actions, and argue that certain standards of transparency and verifiability should be expected from any institution that seeks to involve a statistical analysis in the formation of policies. I will first use Market Share Liability, an established use of statistics, to draw out what responsibilities an institution might have regarding the collection and presentation of statistics. Then, I will propose a system that aims to prevent abusive applications of statistics.

Contents

Dedication

The determined professors that taught
This young academic have not
  Spared effort nor patience
  Despite the frustrations
Of dealing with my oft off thoughts

Acknowledgements

I must thank all of my family and professors for affording me my undergraduate education, but special thanks must be given to my adviser, Professor Kritika Yegnashankaran. Kritika has been an encouraging presence since my very first philosophy class, and I owe much of my interest in philosophy to the many hours I have spent in her office. Kritika's incredible patience, expertise, and good humor has made working on this project a joy. I left every discussion with a better understanding of the problem at hand, and when I got myself confused, or when my writing started to get muddled, I could always count on Kritika to help me straighten things out. Thank you Kritika, for introducing me to philosophy and being such an inspiring presence at Bard.

I must also give special thanks to the other members of my board, Professor Robert Martin, and Professor Marco Dees. Thank you Professor Martin, your enthusiasm for math and philosophy encouraged me to pursue both. The conversations that started in class and overflowed to the sidewalk outside always left me eager for more. Thank you Professor Dees, for guiding me through my first ethics class, and metaphysics class, and enduring my first philosophical flailings in both. Your emphasis on simplicity and clarity has changed my writing for the better.

Thank you Adrian, for listening to my Senior Project woes.

Thank you Leo, Ying, and Alex, for one last semester of good company and conversation.

Chapter 1

Introduction

I am concerned with how we use statistics to justify institutional actions. As our capacity to collect and analyze data grows, it is natural for us to be concerned with the appropriate bounds of data collection and analysis. Concerns about privacy and consent are commonplace, and new proposals for integrating data driven technologies are being proposed both by large companies and governments. In this paper, I argue that providing statistical justifications for institutional actions require specific transparency and verifiability criteria that are entirely separate from concerns over privacy and consent, and propose a system that aims to prevent abusive applications of statistics.

In Chapter 1, I will introduce my motivation and identify the problem at hand. I will do this by reviewing common issues raised with data collection and analysis, and then describing an abuse of statistics that lies outside of these issues. The abuse of statistics I am interested in are attempts by institutions to justify actions by claiming deference to a statistic. I will end with a discussion of why I suspect institutional abuses of statistics deserve special attention over the common worries of accuracy, privacy, and consent.

In Chapter 2, I will introduce a legal case that established a court use of statistics on the market share of a harmful product, and identify exactly what justifies the court in using this statistic to inform legal punishment. The upshot of this examination into an established use of statistics in law will be to show that institutions inherit some responsibilities beyond responsibilities of privacy and security if they expect the public to believe that their policies are informed purely by a statistical analysis.

In Chapter 3, I will introduce a case where an institution attempts to justify a policy using a statistic, and then argue that this justification should not be compelling. By appealing to the inherited responsibilities of institutions discussed in Chapter 2, I will identify what is going wrong with our current view of statistics as justification for policies.

In Chapter 4, I will discuss other cases where the unjustified use of statistics identified in Chapter 3 may appear, such as in employment and police practices.

I will conclude with some comments on avenues for future work, and a short comment on what we should be prepared to see in the immediate future regarding the use of statistics by institutions.

1.1 Problems of Accuracy and Interpretation

The first class of concerns I would like to address are concerns about accuracy and misinterpretation. It is common knowledge that it is easy to mislead using legitimate statistics, and our legal systems have had to guard against misinterpretations in the past. As an example, consider the case of Dr. Meadow and Sally Clark.

> Dr. Meadow was a British Pediatrician interested in Sudden Infant Death Syndrome (SIDS) in the U.K. While he was already known for his work on Munchausen Syndrome by Proxy, he began a run as an expert witness on SIDS, attending to cases of multiple SIDS deaths in families. He reasoned that because the chance of having one SIDS death in a wealthy family with no smokers was 1 in 8,500, the chance of having two SIDS deaths in the same family must be about 1 in 72,000,000, a figure obtained by squaring the 1 in 8,500 figure. Using this statistic, Dr. Meadow testified that Sally Clark, a mother of two deceased infants, was probably responsible for killing her children. Sally Clark was found guilty and sentenced to life in prison, failing her first appeal. Her second appeal was successful, and Sally Clark was released, but Sally Clark would soon die of acute alcohol intoxication after suffering the death of her two children, a miscarriage of justice, and poor treatment at the prison due to inmates and staff believing her to be a child murderer.[1]

This case is often cited as a prime example of an abuse of statistics in the court room. Dr. Meadow made two major mistakes.

[1]Court of Appeal[2000]: Clark, R v [2000] EWCA Crim

The first mistake is assuming that SIDS deaths in a family are independent. An assumption that two events are independent means that the occurrence of one event does not affect the probability of the occurrence of the other event. This is an unsupported assumption in the SIDS case as studies suggest a connection between SIDS and a host of related complications, many of which have a genetic or environmental basis.

The second mistake is not considering that, even if granted that the 1 in 72,000,000 figure was correct, the alternative explanation provided, a double homicide, also has an extremely low *a priori* probability, and it is not clear that the the low probability of one explanation implies the high probability of the other explanation.

This seems like a clear case of an abuse of statistics, but the response to this kind of issue with statistical evidence seems just to be more careful when applying and interpreting statistics. At face value, this issue could be fixed by a more rigorous education in statistics.

Related to issues of Accuracy and interpretation are issues of Practicality, where we are concerned with the feasibility of constructing useful statistical tests. To illustrate this, consider the following excerpt from *Little Brother*, a young adult book by Cory Doctorow, science-fiction author and digital rights advocate.

> If you ever decide to do something as stupid as build an automatic terrorism detector, here‘s a math lesson you need to learn first. It’s called “the paradox of the false positive,” and it‘s a doozy.

> Say you have a new disease, called Super-AIDS. Only one in a million people gets Super-AIDS. You develop a test for Super-AIDS that's 99 percent accurate. I mean, 99 percent of the time, it gives the correct result - true if the subject is infected, and false if the subject is healthy. You give the test to a million people.
>
> One in a million people have Super-AIDS. One in a hundred people that you test will generate a "false positive" – the test will say he has Super-AIDS even though he doesn't. That's what "99 percent accurate" means: one percent wrong.
>
> What's one percent of one million?
>
> $1,000,000/100 = 10,000$
>
> One in a million people has Super-AIDS. If you test a million random people, you'll probably only find one case of real Super-AIDS. But your test won't identify one person as having Super-AIDS. It will identify 10,000 people as having it.
>
> Your 99 percent accurate test will perform with 99.99 percent inaccuracy. [2]

Here, Doctorow is making a practical complaint against statistical tests that aim to identify individuals with a characteristic that has a very low rate of incidence. A terrorism statistical test, due to the extraordinarily low rate of incidence, must have an incredibly high rate of success if the test is to be deployed onto a large population. Again, we are faced with a clear problem for the use of statistics, but once more, the solution to the problem seems clear. The challenge is to be aware of the rate of false positives and account for it when calculating the success of a test.

[2] Cory Doctorow, *Little Brother* (New York: Tom Doherty Associates LLC), 47.

The two problems presented, while genuine problems, already have much discussion on how they limit the use of statistics. For both of these issues, the problem seems to be resolved by simply knowing more about the nature of statistical inferences. There has already been pressure to re-orient our public school math education to put more of an emphasis on the interpretation of statistics, and further investigation into these problems is probably best left to statisticians and our public education system.

1.2 Problems of Privacy and Implementation

Alright, so what of problems with statistics that are accurate and correctly interpreted? The next class of issues I would like to address are issues of privacy and implementation.

Examples of violation of privacy in the name of mass data collection are easy to find, but I will present a particularly sinister application here.

> In 2014, the State Council of the People's Republic of China released a document entitled *Planning Outline for the Construction of a Social Credit System*. In it, China describes a system designed to foster trust and good citizenship amongst Chinese citizens through the use of a credit score. This system would be mandatory for all Chinese citizens by 2020, although the article does not mention specifics on how the government plans to phase in this new credit system. One year after this planning document was published, Chinese mega-corporations Tencent and Alibaba jointly announced the release of Sesame Credit, a credit system in line with the description provided by the

> Chinese Government, with some concerning twists. Unlike the American Fico credit score, Sesame Credit's credit score incorporates information from online purchases and activity on social networks, along with the standard information on loan payments and financial assets. The rewards for high scores are also unusual, affecting a Chinese citizen's eligibility for a pan-European Visa, access to shorter lines at national airports, and travel permits to Singapore. This is all in service to China's stated goal of "establishing the idea of an sincerity culture, and carrying forward sincerity and traditional virtues... its objective is raising the honest mentality and credit levels of the entire society", as the algorithm assigning your score is meant to measure good citizenship by sourcing information about your daily habits.[3]

This proposed social credit system contains all kinds of potential breaches of privacy. Although the social network information used might be public, there is something invasive about using social networks to determine things like personal character. There is also an intuition that the government should not be allowed to access your purchasing history to determine what line you can wait in at an airport. Even if purchasing history turns out to be a reasonably good indicator for good citizenship, the issue does not rest on the accuracy or interpretation of the statistic, but rather what the statistic demands in terms of access to personal life.

Less extreme variants of social network crawling are already in use. In order to simplify the hiring process, many large employers have taken to

[3] *State Council Notice concerning Issuance of the Planning Outline for the Construction of a Social Credit System* (2014-2020), translated by Roger Creemers

running resumes and applications through algorithms in order to weed out the clearly unfit, but some companies have taken it a step further and have almost completely automated their human resources department. The pseudo-taxi company Uber for example, automatically fires employees whose approval rating on their phone app dips below a certain threshold[4]. Today, many admissions procedures review the social media presence of their applicants, and although automated profile crawling is not yet commonplace, technologies that attempt to pull information off of profile pages are well in development. Ideally, these kinds of systems could help remove human bias and speed up the recruitment process, but more difficult problems of implementation and use begin to appear even in these already accepted cases.

Another problem in the usage of automatically generated statistical reports is how users know if the algorithm that is generating the reports is malfunctioning if they have no idea how the algorithm works. As any software engineer will tell you, keeping a program up to date with all the relevant inputs is not a trivial task, and in order to maintain a large-scale system like China's Sesame Credit, we often need to rely on bug reports from users to make sure everything is running as it should. To illustrate:

> Eve and Franny are living under China's Sesame Credit. Eve is very knowledgable about the ins and outs of the system, and can explain in detail how the algorithm is generating the report,

[4]Frank Pasquale. "Judge, Jury and Executioner: the unaccountable algorithm," *Aeon*, August 18, 2015

> and what the report should look like given a brief description of the inputs. Franny on the other hand, knows very little about the mechanics of how the Social Credit System works, and like most other citizens, is only aware of very general rules that affect her score. She knows for instance, that her scores tend to rise when she pays back her debts on time, but there are fluctuations in her score that she cannot explain.

When Franny engages in activity within the Sesame Credit System, she will have a difficult time telling whether or not it is working properly. She may not understand any bounds that the algorithm might have on its inputs, and thus might be misinterpreting its results. Franny's lack of knowledge impairs her ability to reflect on the decisions she might make using this generated statistical report as she cannot be sure that the outputs of the algorithm are appropriate, nor can she be sure that her application of the algorithm is within its bounds. Eve, on the other hand, knows immediately when the algorithm has made a mistake and can critique its efficacy in measuring what it attempts to measure, as Eve is privy to *how and why* the judgement appears.

This is not to say that the users of automatically generated statistical reports must be composed of all Eves, it just means that there must be some Eves within the population that can act as a check on how the report generating algorithm is functioning. This means making the algorithm largely public, allowing motivated individuals to make independent checks on the behavior of the algorithm, allowing the algorithm to improve with suggestions from the Eves of the populace. The existence of Eves also allows Frannys to have someone to appeal to when they believe they have been treated unfairly. When we evaluate report generating systems, we should ensure that enough information is made public to allow for Eves to act as independent checks on the performance of the relevant report generating algorithm. While this still seems like a problem solvable through education, we will be returning to a similar reason when we argue for transparency later in this paper. However, this practical problem has a more complicated counterpart, described by Howell in his paper *Google Morals*.

> Beliefs attained by deference to a statistical report can fail to integrate with the rest of the agent's present beliefs, and can fail to provide the proper ground for new beliefs. This prevents the agent from achieving higher degrees of virtue.[5]

Normally, we expect that judgements on other people to be integrated with other beliefs and behaviors, and this integration can form the foundation on which we synthesize new beliefs. It seems that this ability to develop a sense of whatever is being measured, employability in the modern case and good citizenship in the social credit case, is hindered by using a statistical report without understanding its origins. If I start with deferring to a statistical report, and then later decide I want to make a minor adjustment in my evaluations due to some new experience, I have to scrap

[5] Robert Howell, "Google Morals, Virtue, and the Asymmetry of Deference," *Noûs* 48:3 (2014): 389-415

the entire report and start from scratch, because if I do not understand how the report works, I cannot make the necessary adjustments to accommodate my new point of view. Using automatically generated statistical reports does not allow for judgements to be subtly adjusted over time like we expect judgements of mind to be.

Here we hit a problem that is closer in kind to our central problem of interest. This is a problem that is not a practical concern, cannot be fixed by statisticians, has little to do with the privacy or consent of whoever is being examined by the statistic, but still points to bad consequences of a specific abuse of statistics. Our problem of interest, which will be presented in this next section, distinguishes itself from commonly discussed problems about statistics in similar ways.

1.3 Our Problem

I am specifically concerned with institutions that attempt to justify their policies and actions through the use of statistics. As an example, consider the CVS security tag case.

> In 2006, CVS fell under investigation when customers noticed that in a selection of CVS locations, only hair products marketed towards black women had an attached security tag. When accused of racism, CVS responded that they merely make a practice of affixing security tags to products stolen most often. Presumably, somewhere in CVS offices, there exists a database that keeps track of products that disappear off the shelves, and

> it is this mere deference to a statistical database that produces this apparently horrifically racist act. Is CVS correct in thinking that they did no wrong?[6]

I am bothered by this kind of deployment of statistics, but it is somewhat difficult to place why. After all, my discomfort persists even if I suppose the statistic is perfectly accurate, correctly interpreted, and no employee holds a specific racist attitude. Additionally, there does not seem to be anything that interferes with privacy or consent in the collection of the statistic in question, as the statistic is not about customers, and rate of theft seems like an appropriate metric to base the distribution of security tags. However, there are a few things that hint that there is something going wrong with this use of statistics.

First, even though we cannot definitively conclude that there are any individuals that hold racist attitudes in CVS, we can still accuse the CVS institution as being a racist institution. If we allow for institutions to be racist at all, it becomes immediately apparent that a specific racist belief or attitude in the conventional sense is not necessary to level a complaint about racism, as institutions do not hold beliefs or attitudes in the same way individuals do.

Second, it has become unfortunately common to use statistics to try and end difficult conversations. Too often, questions on whether or not racial profiling is permissible, whether or not to place travel bans on cer-

[6] "CVS Hair Care Products," last modified June 22, 2014, www.snopes.com/racial/business/cvs.asp

tain countries, and whether or not certain races should be placed under specific surveillance are met with simple relays of statistics of the form "such and such class of people are responsible for a high percent of such and such harm" or "This decision cannot be a bias incident, as it was based off of such and such statistic". This use of statistics is difficult to meet, as it seems that statistical information is a near inarguable empirical fact. However, hidden behind the spouting of these statistics is a dangerous assumption that deference to a statistic is inherently unproblematic, an assumption that I will go on to challenge later.

So, what exactly is the role of statistics in justifying institutional actions? In what cases is deferring to a statistic appropriate, and when does it constitute an abuse? I will begin our investigation by looking into some established legal philosophy for some insights.

Chapter 2

Statistics and Liability

Consider the following hypothetical.

> A horrible bus accident has occurred, and the Blue Bus Company is being sued by the victim. This accident occurs in two worlds exactly alike except:
>
> In the first world, the victim reports that the bus she was hit by had the Blue Bus Company logo. The court knows that eye-witness accounts are fallible, and that they are accurate only about 70% of the time.
>
> In the second world, the victim never sees the bus that they were hit by, but it is known that 70% of the buses in the city are from the Blue Bus Company.
>
> Is the eye-witness evidence presented in the first world just as good as the statistical evidence presented in the second?

Enoch, Spectre, and Fisher provide an account of statistical evidence that identifies two properties that eyewitness evidence has that statistical methods do not have.[7] The first is that eyewitness evidence is sensitive to counterfactuals, and the second is that eyewitness evidence creates an incentive system that discourages the breaking of laws. I address sensitivity first.

2.1 Sensitivity

Measuring something generally requires an instrument that is sensitive to whatever it is being measured. For instance, a thermometer is sensitive to temperature because it changes in accordance with the temperature. When I see a thermometer that reads 30 degrees Celsius, I know that if it was not 30 degrees Celsius then the thermometer would not read 30 degrees Celsius, and I know that if the thermometer did not read 30 degrees Celsius, then it is not 30 degrees Celsius. We can test if an instrument is sensitive to an outcome x by checking these two statements.

- If the instrument reads x, then it is x.
- If the instrument does not read x, then it is not x.

For our thermometer, this would mean:

- If the thermometer reads $30°C$, then it is $30°C$.
- If the thermometer does not read $30°C$, then it is not $30°C$.

[7] David Enoch, Levi Spectre, and Talia Fisher, "Statistical Evidence, Sensitivity, and the Legal Value of Knowledge," *Philosophy and Public Affairs* 40 (2012)

We can construct an analogous pair of statements when considering evidence. We might say something like "In order to use a piece of evidence to determine the guilt of some party, the piece of evidence must be sensitive to whether or not the party is guilty." In this case, the instrument is the eye-witness account, and the target measurement is the bus company responsible for the accident. However, we cannot directly port over our original two statements, as it is not true that If an eye-witness makes a testimony that the blue bus company is responsible, then it is the case that the blue bus company is responsible. As our eye-witness is very fallible, we need a weaker condition that keeps the separation. We can instead focus on whether the quality of the evidence changes when considering a counterfactual. For instance, in the eye-witness case, we can ask if it would be likely for the eye-witness account to change if the blue bus company did not cause the accident. Here, the answer seems to be yes. If the blue bus company did not cause the accident, then the eye-witness would probably not testify that they did. However, in the case of statistical evidence, if the blue bus company did not cause the accident, the evidence would not change. It would still be true that the Blue Bus Company owns 70% of the busses in the city. And so, because this statistical evidence is not sensitive to counterfactuals, we have reason to prefer the eye-witness evidence over statistical evidence, even though they are both equally fallible.

2.2 Incentives

Related to Enoch et al.'s concerns about sensitivity is their incentives account. To illustrate this account, consider the following thought experiment proposed by Enoch et al.

> Someone named Alice thinking about scalping tickets to a show. She is standing outside of the gates, wondering if she should begin selling her tickets for profit. She looks around, and observes that 80% of the people around her are scalping tickets. Like our Blue Bus case there are two possible worlds, one in which the legal system uses eyewitness evidence, and one in which the legal system uses statistical evidence.
>
> In the first world, where the legal system uses eyewitness evidence, Alice has an incentive to not scalp the tickets, as if she does, someone may see her and provide eyewitness evidence, and if she does not, there will be no eyewitness to see her scalping tickets. In the second case, where the legal system relies on statistical methods, Alice has no incentive to not scalp tickets, as the court will conclude that she is 80% likely to be scalping tickets, from observing the proportion of caught scalpers to non-scalpers. So statistical evidence should be viewed with caution as it does not provide an incentive system that discourages breaking the law. [8]

Setting aside the concern that the functions of law may or may not include proper incentive systems, this account seems satisfactory in small cases like ticket scalping, but in larger cases, such as accusations of terrorism, these legal incentive systems appear to be very weak. For example,

[8] Ibid.

we can imagine that a person about to board a bus would not be inspired to bomb the bus by the sheer fact that they would be accused for it anyway under a legal system that uses statistical methods. It should be noted that many contentious statistical methods in law, like racial profiling, are in use in the hopes to catch extreme crimes. For cases like this, it seems the incentive account is incomplete, and we need something a little stronger to justify ignoring or accepting statistical methods in our legal proceedings.

Enoch's diagnosis of the ticket scalping example works because the scalping of tickets is a wrong of exploitation for personal gain. The criminal in this case has a clear benefit in doing the crime and need only to check whether or not they would be caught. Crimes of exploitation, or intentional harm, need laws that are sensitive to the particulars for the exact reason that Enoch provides. Without sensitivity, there is no incentive to avoid doing the crime. But we might go further and say that the lack of this incentive system justifies the crime should it happen. Consider the IRS taking a look at tax forms, and suppose that it is known that the majority of people overstate their charitable contributions by about $500. The IRS then decides to deduct $500 from all statements of charitable donations. The worry here is not that the IRS will be less accurate, but instead that the IRS has now:

1. Justified the tax payers who overstate their contributions
2. Punished those that do not overstate their contributions

This issue still stands should we consider a probabalistic case. Suppose we determine that 70% of the population overstate their charitable contributions. Any way we split the cost, the moment the IRS declares a flat deduction to compensate for this overstatement, overstating by that much becomes justified, and the strategic course of action.

2.3 Statistical Confidence

Schmalbeck gives us a separate account of why statistical evidence might leave us uncomfortable.[9] Can we explain our discomfort by appealing to statistical confidence? We can imagine a potato inspector looking over a shipment of 1,000 potatoes. If the passing standard is that no more than 5% of the potatoes are too small, and our inspector opens a box of 50 potatoes and finds 4 potatoes are too small, the best the inspector can do is infer that 8% of the potatoes in the shipment are defective. While the judgement is justified, we must attend to how confident we are that our inference is a good estimation of the actual distribution of small potatoes. In this case we could just take a larger sample, but for our blue bus company, sample size cannot be an issue, as we are given that we know what company all the busses in the city belong to. However, we can still pin some of the blame on a shaky level of confidence we have in our statistic. Instead of separating out distributing liability using the share of busses or share of prior crimes, factors such as the distribution of busses and share of prior

[9] Richard Schmalbeck, "The Trouble with Statistical Evidence," *Law & Contemporary Problems* 49 (1986)

crimes can just be said to increase our level of confidence in our statistic. If we can explain our discomfort with using statistics in law by our lack of attention to confidence, we should investigate whether or not we can identify some confidence level that would ease our worries.

While it seems true that Statistical Confidence is a relevant concern with statistical evidence, Enoch et al.'s concerns and Schmalbeck's concerns seem to be of different kinds. Issues of Sensitivity and Incentives appear robust even when we assume maximum statistical confidence. Just as before, it seems like Schmalbeck's concern of statistical confidence lies in a practical concern that can simply be fixed using more accurate measurement methods. It does not point to a concern that is inherent in the idea of applying a statistical test.

We have already reviewed the Enoch et al. account and Schmalbeck's account of why we should prefer eye-witness evidence, but we are now interested in whether or not statistical evidence can be used as part of the sentencing process. While it seems like we have reason to be very suspicious about the use of statistical evidence, there are more moderate answers we can consider. So far, we have attended to accounts of statistical evidence that consider its influence on who should be charged, but in the following sections, we will see a use for statistical evidence used directly in the sentencing process.

2.4 Market Share Liability

One way to use statistics to resolve the bus problem is to outline a sort of distributed liability that holds the blue bus company responsible only for a portion of the claimed damages. A legal doctrine already exists for portioning out liability, but this doctrine currently only covers product liability.

The doctrine we are interested in extending is the doctrine of Market Share Liability, established by Sindell v. Abbott Laboratories. In Sindell, a woman who developed a cancer related to a drug her mother took while she was pregnant sued eleven drug companies on behalf her herself and other women in similar situations. The drug, DES, was marketed to pregnant women to help avoid miscarriages. When it was found to be linked with cancer in children, the FDA ordered that the drug be labelled as experimental, with additional warnings for pregnant women. The defendants continued to advertise the drug without restriction, against the orders of the FDA.[10]

The notable portion of the case was a cause of action that stated the defendants were jointly liable regardless of which of the defendants manufactured the brand of DES that the plaintiff's mother consumed. Due to joint marketing campaigns and pharmaceutical practices at the time, the specific brand that was consumed could not be identified, and so any joint

[10]Sindell v. Abott Laboratories, 449 U.S. 912 (1980).

liability that would be established could not depend on knowledge of the actual manufacturer.

In the end, the defendants were held responsible for the damages in proportion to the market share they held on DES. The suggested requirements for this kind of liability was as follows:

1. There existed an insufficient, industry-wide standard of safety as to the manufacture of the product.
2. Plaintiff is not at fault for the absence of evidence identifying the causative agent but, rather, this absence of proof is due to defendant's conduct.
3. A generically similar defective product was manufactured by all the defendants.
4. Plaintiff's injury was caused by this defect.
5. Defendants owed a duty to the class of which the plaintiff was a member.
6. There is clear and convincing evidence that plaintiff's injury was caused by a product made by one of the defendants. For example, the joined defendants accounted for a high percentage of such defective products on the market at the time of plaintiff's injury.
7. All defendants were tortfeasors[11] [12]

That fifth condition, the defendants owing a duty to the class of which the plaintiff was a member, will be the condition of interest later in our

[11]Ibid.
[12]A *tortfeasor* is one who commits a tort. A *tort* is a civil wrong

generalization of Market Share Liability, as many of the the other conditions are specific to a product defect.

There are some difficulties with this kind of distributed liability. With the establishment of any new legal doctrine, we must be careful to consider possible avenues for exploitation. It is immediately clear that a strong method of determining the "market" for other kinds of statistical evidence is critical. As an example, for our bus case, we observe that the blue bus company owns 80% of the busses in the city, but owns only 40% of the busses that have routes running through the site of the accident. If we have a competing bus company, the red bus company that owns all other busses in the city, then the blue bus company is motivated to present the more constrained statistic, and the red bus company is motivated to present the broader statistic. If we allow for each defendant to select the statistic that presents them with the smallest share of liability, we would be judging each defendant with different standards, and would be unable to recover a significant portion of the damages. In this instance we would only be able to recover 60% of the damages, 40% from the blue bus company, presenting the more constrained statistic, and 20% from the red bus company, presenting the broader statistic. In the interest of fair judgement and to make it possible to recoup a significant share of the claimed damages, we must add an additional condition that sample in the statistics used must be the same.

2.5 Problems With the FDA

There are some practical blockages to any account of distributed liability. For instance, in Sindell, the FDA acted appropriately by judging that DES be labelled as experimental, and it is clear that the defendants in the case were at the very least guilty of ignoring the commands of the FDA. Because the FDA acted appropriately, it seems easy to limit the liability to the companies that did not follow the FDA guidelines. However, suppose we have a case where the FDA acts appropriately, the drug companies act appropriately, but damages are still incurred. Our example for this case will be the Reye's Syndrome link with Aspirin.

Aspirin, being one of the most commonly used medications in the world, was considered well understood, and relatively safe. However, in the 1980's a series of epidemiological studies demonstrated a link between children taking Aspirin for specific illnesses, and Reye's Syndrome, an often fatal kind of swelling of the brain. The FDA acted quickly to demand warnings, and drug companies complied. Although the FDA worked as it should, and the drug companies did indeed follow orders, who, if anybody, should be held responsible for the damages associated with having a child afflicted with Reye's Syndrome?

Two practical barriers affect our notion of distributed liability. The first is Sovereign Immunity, which states that the government and governing entities cannot be sued unless the federal government waives its immunity. The Federal Tort Claims Act[13] gives private parties the right to sue the United States if a tort is committed by a person acting on behalf of the government, but the Act has, written within it, a clause preventing lawsuits against the FDA that implicate the drug approval process. So, even if a case could be made that an employee at the FDA committed a tort while acting on behalf of the government by approving a defective drug, the FDA is immune from all such lawsuits.

The second barrier is Wyeth v. Levine, which judged that FDA clearance of a medication does not protect against liability claims under state labelling laws.[14] This stops us from trying to move liability up the chain to the drug approval system, as a drug passing federal clearance does not grant the drug company freedom from liability.

There seems to be a practical problem of what to do with the missing liability if we find that the FDA or some other legal entity with protected status is partially at fault for whatever harm we are examining. While these considerations might stop us from insisting that our account of appropriate uses of statistics be integrated directly into the legal system, we can, and will later in the paper, argue that there is still a responsibility to transparency and verifiability.

[13] Federal Tort Claims Act, 28 U.S.C. Pt.VI Ch.171 (1946).
[14] Wyeth v. Levine, 555 U.S. 555 (2009).

2.6 Sindell and the Bus Company

Our comfort with claiming that the blue bus company was an analogous case to the Sindell case without false advertising is that we noted that passing FDA regulations does not grant legal protection from litigation, as drug companies are still responsible for determining the effects of the drugs that they produce. However, this move was too quick, as although the bus companies are all still responsible to drive safely, they do not have a responsibility to demonstrate that their driving did not cause a specific accident. To illustrate this difference, we can examine several adjustments of the Sindell Case. A variant where none of the companies falsely advertised their product, and a variant where there was no error in production or testing the medicine.

Our original Sindell case seems to be a clear cut case where distributed liability is a good option. The fact that all of the defendants are known to have committed false advertising and did not follow FDA guidelines seems like a strong enough reason to say that the defendants are all responsible for some portion of the damages. However, the main consideration when thinking about justified distributive liability is not whether or not all defendants are tortfeasors or if they knew their product could cause harm, but whether or not the company is *responsible for collecting evidence that would demonstrate that they could not have contributed to the harm.*

After we know that one of the defendants caused a harm, that harm is causally linked with some action or inaction that the defendant has a responsibility to keep to track of. Then, if any of the defendants can produce the required records that show that they were not responsible for that harm, they are ejected from the liability pool. The remaining defendants that are unable to produce the relevant records are subject to distributed liability. In the case where all defendants produce the required records and pass, the court was incorrect in their evaluation that the harm was caused by that action or inaction. The court must then review the cause of the damage, and the pool of defendants cannot be litigated against. For instance, in the variant of the Sindell case where the defendants do not falsely advertise and do not know if the drugs they produced were the harmful drug, we are still justified in including potential non-producers of the harmful drug in the pool of defendants because drug companies should have records of what kinds of drugs they produced.

Here, the defendants have a responsibility to provide evidence that is available for public review.

	All defendants are tortfeasors	Knows that their product can cause harm	Defendants should have exonerating evidence	Is distributed liability justified?
Original Sindell	Yes	Yes	Yes	Yes
No False Advertising	No	Yes	Yes	Yes
No Production or Testing Error	No	No	Yes	Yes
Bus Case	No	No	No	No

To be a true parallel to our bus case, we should have a pool of defendants that are being included in the pool of defendants due to some feature they share that they do not have a responsibility to track. For instance, we can imagine that Sindell, instead of being able to identify the drug, instead remembers that she purchased it on the third shelf of some convenience store, and so is suing all drug companies that happened to have drugs on that shelf. We see here that it seems unreasonable for drug companies to keep track of what shelf their medication is in, and furthermore, their involvement in this case seems purely predicated on the shelving practices of the convenience store. We might still have an intuition that these drug companies should still be included in the pool. We can imagine a harsher case. Imagine that Sindell had suffered medical complications from consuming food from a farmer's market. Sindell can only identify that she purchased the food at a time between 3-4 pm, and she is suing all individuals that sold food in between those times. Defendants are no longer drug companies but Grandma Jones, who sold her spinach that she grew in her backyard, Fred, an unassuming businessman with a hobby for horticulture, and Rachel, a young college student selling decorative cacti to help with her student debt. Are we willing to saddle these people with the responsibility of paying for Sindell's bills?

The answer hinges on whether or not the court decides that Grandma Jones, Fred, and Rachel are responsible for testing their food for harmful pathogens, which in this case seems unreasonable, as they do not pose as large a threat to public health i comparison to a restaurant, and they do not undergo the regular health and safety checks required by local law. Below, we present a case that illustrates what we demand.

2.7 Restaurants and Dinner Parties

Suppose Alice and Bob are each very hungry and decide to eat diner at three different places. Alice chooses to three different restaurants called restaurant 1, restaurant 2, and restaurant 3. Bob chooses to attend three different dinner parties hosted by friends, called party 1, party 2, and party 3. Alice and Bob, after their night of eating three dinners, comes down with a terrible food borne illness, and are looking to litigate against the restaurants and the hosts of the dinner parties respectively. Neither Alice nor Bob can pinpoint which restaurant or party caused the illness, and so they attempt to take all of them to court. The court determines that the cause of Alice and Bob's illness is a pathogen that is active when meat is not kept at the proper temperature.

In the case of Alice, it is known that restaurants are required to keep daily records of their refrigerator temperatures, a health and safety requirement in some states, and so the court orders these records from the restaurants. Suppose restaurant 1 produces the records, which state that proper fridge temperatures were maintained, but restaurants 2 and 3 could not produce the required records. Then restaurant 1 is removed from the pool of defendants, and restaurants 2 and 3 are subject to distributed liability. Each of restaurants 2 and 3 are responsible for 50% of the damages that Alice is claiming, even if we have no idea which restaurant is the true source of the pathogen.

In the case of Bob, it is not at all required for dinner party hosts to keep records of their fridge temperatures, and so Bob cannot litigate against his hosts. The case is dropped and none of the hosts are held responsible because Bob cannot identify which of them are responsible, and its unreasonable to require that the defendants have evidence that would determine their eligibility to b e included in the pool of defendants. It is unreasonable to require dinner party hosts to keep regular temperature records because dinner party hosts hold significantly less capacity to harm the public. Individual dinner party hosts do not hold the kind of institutional power an established public restaurant would.

2.8 Distributable Forms of Punishment

It appears that only certain kinds of sentences can be dealt with by distributive liability. Going back to our shotgun case, if Alice and Bob both shot a shotgun shell with 50 pellets at Carl, and Carl died from a single bullet, it seems inappropriate to argue that Alice and Bob should evenly split the sentence. Taken to the extreme, all one would need to do to effectively get away with murder is to invite as many friends over as possible, hand all of them shotguns, fire at Carl, and then enjoy $\frac{1}{n}$th of the share of the punishment, where n is the number of friends you invite. From this, we might suspect that only financial liability is divisible, or that the distinction lies in the civil/criminal distinction.

However, we can refine the method of distributing liability so that we do not need to appeal to either financial liability, or the civil/criminal distinction. A more nuanced approach to the shotgun case is to claim that we should subtract the sentencing for a successful murder from an attempted murder, and then distribute that figure among the shotgun wielders. This method seems to handle non-financial sentences more appropriately than the blind split. Additionally, if we grant that we can distribute liability in this criminal murder case, and in the civil Sindell case, we cannot appeal to the civil/criminal distinction to decide when distributed liability is appropriate. We are justified here in dividing up the sentencing because there is no question that everyone is guilty of attempted murder. What is to be determined is how the statistic informs who fired the fatal shot. Imagine for instance that the sentence for attempted murder is 20 years, but the sentence for a successful murder is 26 years, and Alice and Bob are the only shotgun wielders with an equal number of pellets fired from each gun. It does not seem implausible that the proper response is say that Alice and Bob should both get 23 years. However, this runs counter to our restaurant/dinner party distinction! The shotgun case places Alice and Bob in a dinner party case, as we do not expect that Alice or Bob should collect statistics on where their pellets went after they fired. If we are to simply extend our earlier proposal, then using distributed liability is unjustified, and Alice and Bob should not split the additional six years. However, the function of the restaurant/dinner party distinction is to clear up the fuzziness of proposing a cause, and then seeing who is potentially responsible for the proposed cause. Then in the shotgun case, there is no mystery of the cause of the crime, and so Alice and Bob already volunteer themselves to distributed liability when they took aim with their shotguns.

2.9 Our Account

Here, we are ready to present a full account of the generalization of Market Share liability:

1. First, we have a defendant who is claiming action against a group of defendants, from which the defendant cannot identify who specifi-

cally committed the tort.

2. Then, the court decides on some cause of the damage.

3. The court then investigates whether or not the defendants have a responsibility to keep records of actions that would demonstrate that they were not the cause.

4. If such a responsibility exists, all the defendants that could not produce the expected records or fail to have the records demonstrate freedom from liability can be subject to market share liability. If no such responsibility exists, then distributive liability is unjustified.

With this method, we can solve our case of the Blue Bus Company by considering the cause of damage.

One outcome might have the court decide that the cause of the harm was a mechanical failure of the bus. The court discovers that all bus companies must submit daily inspections on their vehicles and asks the group of defendants, which includes the Blue Bus Company, to produce the reports. If the Blue Bus Company can produce the report, and the report states that all their buses were in good working order, then the Blue Bus Company is ejected from the pool of defendants and is not responsible for any damages. If they fail to produce the reports, or the reports show that some of the buses had the specified mechanical failure, the Blue Bus Company cannot prove that they could not have been responsible, and so distributed liability is applied.

Another outcome might have the court decide that the cause of the harm was an inattentive driver. After investigation, the court finds that bus companies are not required to subject their employees to the kind of supervision that detects attentiveness, so the case is dropped and the Blue Bus Company pays nothing. Another option is to have distributed liability as an option only when the defendants have been operating under some regulatory body that dictate exactly what must be tracked. What this regulatory body finds is reasonable to demand is left up to the regulatory body. Just as we have health and safety regulations, we can propose statistical analysis regulations. I will discuss this kind of regulation in more detail in the next chapter.

2.10 Self-incrimination

An immediate objection to the restaurant and dinner party distinction is that in the restaurant case, it seems like the restaurant is being required to self-incriminate, or put themselves in danger of persecution. However, we can simply move when we check the records for conformity to an already established practice of health and safety inspections, and records that are already relayed to regulatory bodies. There is an open question on whether self-incrimination extends to institutional policies, but just as it has become popular to demand that police wear body cameras, it might be reasonable to institute a practice of requiring each refrigerator in restaurants to contain a record-keeping thermometer. When the health inspector comes along, just as they would ensure that every fire extinguisher is in place and being checked daily for pressure, they would also ensure that the appropriate thermometer is in place and being recorded.

What we have done here is identify some additional criterion that is required when attempting to use statistical evidence to justify an institutional action in reference to a regulatory principle. In the restaurant and dinner party case, we say that in order for us to use market share in order to justify distributed sentencing, the defendants must have a responsibility, or duty, to the plaintiff to provide the relevant statistic. In the restaurant case, it was refrigerator temperatures, and in the bus case, it might be mechanical checks, or anything else that we might judge a bus company responsible for.

2.11 Limits of Marketshare Liability

We may be comfortable with making judgements using marketshare liability on companies, but making judgements on individuals seems intuitively more difficult.

Dr. Thomson, in her paper *Liability and Individualized Evidence*, proposes another related thought experiment. Suppose Alice and Bob both do not like Carl, and both shoot Carl with a shotgun at the same time. Carl is killed due to the action of a single pellet, and we are unable to determine whose shotgun fired the pellet. After some investigation, it is discovered that Alice's shotgun fired a shell with 95 pellets, and Bob's shotgun fired a shell with only 5. They are both clearly guilty of attempted murder, but who is guilty of murder?[15]

Our distinction between restaurants and dinner parties doesn't seem to apply here, as neither the defendants nor the victim are institutions that might carry a responsibility to track a specific statistic. If we are to claim that distributing the sentencing based on the probability that one of their bullets was the lethal bullet, we must justify that use of statistics with something other than the process I have described here.

to make things more difficult, we can consider scenarios in which there is no institution that has done wrong, and it is not the case that all accused are guilty of even an attempt at harm. For instance, consider the case where ten people are in a bus, and someone drops a ball filled with toxic gas. The ball rolls about the bus and bursts open, and all ten people need expensive medical attention. We know that one of the ten persons on the bus is responsible, but we wouldn't claim that each person is responsible for 10% of the damages.

But in this paper I am not too concerned with the application of statistics to individuals. The useful tool that comes out of the discussion of Marketshare Liability is actually the restaurant and dinner party distinc-

[15] Judith Thomson, "Liability and Individualized Evidence," *Law & Contemporary Problems* 49 (1986)

tion, which shifts responsibility for recording and producing statistics onto institutions.

Chapter 3

Statistics as Justification

I will move away from discussing distributed liability for a moment while we attend to the CVS case introduced the introduction. While I briefly mentioned some intuitions about the CVS case that pointed to some discomfort in the Introduction, here, I will outline the specific problems that come with trying to use statistics as a justification for institutional actions. The end goal is to make it very difficult for an institution to abuse statistics to deflect blame for things like racial bias, and provide a system where the public can be confident that institutions are being honest about their policies that they claim are being motivated purely by statistical analysis.

3.1 The Masking Problem

Statistics, because of their perceived status as objective and factful, are easy tools to divert suspicions. To help illustrate the potential harm in using statistics to justify actions, I will offer some variations on the CVS case presented in the introduction, copied here for convenience.

> In 2006, CVS fell under investigation when customers noticed that in a selection of CVS locations, only hair products marketed towards black women had an attached security tag. When accused of racism, CVS responded that they merely make a practice of affixing security tags to products stolen most often.

The most obvious worry is whether or not CVS was being truthful about its tracking. It seems plausible that in order to cover up the racially motivated actions done by a few managers, CVS just cited a database that did not exist. As CVS never released the data that would tell them which items were most often stolen, we are forced to just trust them that the data exists, and was indeed the reason behind the security taggings. Here, the citing of the statistic just helps mask racial motivations.

The problem is not dissolved if we suppose that the database exists and is accurate. We can imagine an adjustment to the CVS case where managers of individual locations are given a choice as to whether or not to use the database to tag products. A racist manager could easily base their decision on which protocol disadvantages a race more. For instance, when selecting which protocol to use when tagging products, say a protocol where random items are tagged and a protocol where the database is used, a manager might first check whether or not either protocol disadvantages black customers and base their decision on that determination. When accused of racism, the manager denies that they chose the protocol based on race and instead says that they were purely motivated by trying to keep the theft of objects at a minimum. Under this system, once again, the deployment of the statistic serves to mask racially charged motivations.

I will use the term *masking* to refer to cases where an institutions policy that is justified through deference to a statistic has the effect of excusing actions that are born our of other motivations.

With these adjustments, we can see that claiming deference to a statistic, even if the statistic is accurately sourced and correctly interpreted, is not enough to justify actions. However, the core problem still seems to rely on some individual having a racist attitude. Can an institution abuse a statistic even when none of its constituents hold a specific objectionable attitude?

3.2 Bias in Institutions

Attending to bias in institutions has already had attention from our legal systems. In 1961, President Kennedy signed an executive order that required government contractors to "take affirmative action to ensure that applicants are employed and that employees are treated during employment without regard to their race, creed, color, or national origin."[16] A few years later, the Equal Opportunity Employment Commission (EEOC) was formed, which attended to cases of workplace discrimination for some protected classes. Of course, as one might expect, these legal measures to stem discriminatory practices ran into some immediate practical problems.

The immediate problem recognized was that mentioned in the previous section. What is required to determine an institution guilty of discrimination? If we require that there must be an attitude held by some employee, it seems like there is no need to attend to purely institutional cases of discrimination. A suit can be filed against the specific individual accused of holding this attitude, and the repercussions for the institutions would fall out of attempting to undo whatever policies have been implemented by that specific person.

While masking has already been recognized by a few philosophers, the prevailing common view seems to be that wrongs of bias reduce to attitudes of specific individuals. In *Judge, Jury, and Executioner, the Unaccountable Algorithm*, Frank Pasquale points out that states have already passed laws aimed to restrict masking in other cases.

> Consider a variable that seems, on its face, less charged: months since last job. Such data could aid employers who favour work-

[16] Exec. Order. No. 10925, 26 FR 1977 (March 8, 1961), https://www.archives.gov/federal-register/executive-orders/1961-kennedy.html.

> ers quickly moving from job to job — or discriminate against those who needed time off to recover from an illness. Worried about the potentially unfair impact of such considerations, some jurisdictions have forbidden employers from posting "help wanted" ads telling the unemployed not to apply. That is a commendable policy step — but whatever its merits, what teeth will it have if employers never see CVs excluded by an algorithm that blackballs those whose latest entry is more than a few months old? Big data can easily turn into a sophisticated tool for deepening already prevalent forms of unfair disadvantage.[17]

We can recognize similar problems with other variables in other systems. Hiring procedures that take a look at zip codes, scholarship funds, and financial background raise worries that they are essentially constructing a race test using metrics that correlate closely with race. The Chinese Social Credit System had proposals to track online purchases, which might seem innocent until we consider what might happen if purchasing supplies for infants or purchasing large amounts of imported goods might be used to discriminate against pregnant women or foreign born citizens. While any one of these inputs might seem to loosely correlated for a good statistical test, a large scale algorithm operating in "big data" may be the perfect masking mechanism for protected classes.

At the very least, institutions can be accused of being structured in such a way that makes it very easy to hide discriminatory procedures. If we are

[17] Frank Pasquale. "Judge, Jury and Executioner: the unaccountable algorithm," *Aeon*, August 18, 2015

to take seriously our legislation designed to minimize discrimination, we must be careful about our attitude towards statistical justification. The use of statistics cannot be seen as a thing of convenience, nor can the deference to a statistic be the end of a conversation about discrimination. Large institutions under regulations for discrimination should be designed to have the fewest possible avenues of exploitation, which suggest that we need additional criterion for appropriate uses of statistics.

3.3 Habituation and Normalization

In the following three sections, I will give some wrongs in using statistical evidence that surface without any objectionable attitude held by any constituent of an institution.

A problem arises when the reason for institutional policies is hidden. We can imagine an adjustment of the CVS case where no employee in CVS is racist, and all the store managers choose to use the database of stolen items purely to reduce the number of stolen items. With this presumption, any problem of expressing bias cannot be reduced to the bias of individuals, as no individual in CVS has an objectionable bias. When customers enter the store, they notice that the products that are tagged are marketed towards black people. Just having a policy that tags products in this way opens up avenues for habituation and normalization.

Habituation is the process by which an agent decreases or increases a response to a stimulus through repeated applications of that stimulus. Normalization occurs when previously unacceptable behaviors become commonly accepted.

In our hypothetical CVS case, we can imagine that after this policy is implemented, customers are initially shocked and demand answers. The managers reply that their store tags in this way in accordance with a loss prevention database, which does not consider any racial marketing that a product has. The customer then goes on their way, and in every CVS store they walk into, they notice that the products are tagged in this fashion, and remember that CVS has this database. What will happen when these customers encounter incidents of explicit racism? Eventually, seeing products marketed towards black people selectively tagged will not prompt these habituated customers to seek explanations, or they will assume that the explanation is that the store has a similar database, even when they don't. Widespread use of this policy desensitizes the population to responding to signs of bias, thereby assisting in masking racism in institutions that *do* contain individuals that hold objectionable attitudes.

With Habituation, comes worries of Normalization. As people respond less and less to these practices, over time, the behavior can become the new norm, even without the original justification. We can imagine another shop owner, too small to maintain their own database, simply copying CVS practices as they too want to reduce the number of thefts in their store. With customers being habituated out of responding to apparently racist practices, we risk masking the explicit bias of other institutions, and normalizing the behavior into acceptable without reference to the statistic that bore the behavior in the first place. There are a selection of practical reasons to worry over these normalized behaviors. First, separating the statistic from the behavior makes it impossible to check whether or not a manager is explicitly racist, or is copying the now normal behavior of an institution like CVS. Second, it provides traction for existing prejudices. It is already popular for far-right nationalists to recruit new members by spouting statistics on race, and normalizing this behavior might make this transition much easier. Third, separating the behavior from the statistic renders the behavior insensitive to change. We might expect CVS's in different locations to have different products that are most often stolen, or businesses of different kinds to have different rates of theft. In addition, these statistics may change over time, rendering the normalized behavior obsolete, and the store owner no method of checking whether or not the policy is actually serving its intended purpose.

These problems, habituating and normalizing actions, masking racist institutions, and justifying the rhetoric of racist institutions, occur without any intention on the part of the institution that implements this policy.

3.4 Other Accounts of Racism

There are other accounts of Racism that can also be applied independent of any individual holding racist attitudes.

- The *Responsibility Account* maintains that racial profiling against races that face background injustice is more problematic if the higher offender rate in the profiled group is caused by social injustices for which other groups are responsible.
- The *Expressive Harm Account* maintains that racial profiling is problematic because it makes background injustice vivid and thereby causes the profiled to feel resentment.
- The *Humiliation Account* maintains that racial profiling is problematic as it places individuals in a situation where they cannot prevent appearing to onlookers in a demeaning way.[18]

These accounts do well in describing the harms of racial profiling, but do not generalize well to cases beyond racism. I aim to give an account of why using statistics as a justification in general can be problematic without specific criterion. Specific classes of statistics as justification, like racial profiling, can have additional reasons why they might be objectionable, but our account will hit more than protected classes such as race.

We can imagine, for instance, some warehouse company that does not want to spend extra money purchasing step ladders, and so they try to avoid hiring short people. The responsibility and expressive harm based

[18]Bou-Habib, Paul. "Racial Profiling and Background Injustice." *The Journal of Ethics* 15, no. 1/2 (2011): 33-46. http://www.jstor.org/stable/41486899.

account fails because there is (arguably) no background injustice for short people. The humiliation account seems to fail if all hiring is done in private, and the company never divulges the reason why applicants are not hired. There still seems to be something wrong with this kind of hiring, and so we should look for accounts of policy making that stop such abuses even without the context of background injustice or matters of appearances.

3.5 What Must be Done

Before I show the criterion for statistical justification, I will generalize the CVS case. Any time some institution institutes a policy that is suspected to be born from some bias, and the institution justifies the policy by referencing some statistic, the previously outlined problems are imminent. In order to properly avoid these issues, we must look back at our restaurant and dinner party distinction mentioned earlier.

To review, we imagine Alice has come down with a terrible case of some food-borne illness that is caused by improperly refrigerated foods, and we know that one of three restaurants is potentially responsible. Alice accuses all three. As the restaurants are required to record daily refrigerator temperatures to submit to a health and safety regulatory body, we can check if any of the restaurants had refrigerators that were not up to standard. If any of the three restaurants show that they indeed had the proper refrigerator temperature, they are free to go, but whichever restaurants fail to provide the records or fail to meet standards, we are justified in using a form of distributed liability amongst them.

The reason why distributed liability is appropriate in this setting is because it has been determined that in contrast to a dinner party, a restaurant requires certain licensing that demands compliance with some health and safety guidelines. To fail to produce the required report is to open yourself up to accusations of breaches in health and safety. These guidelines protect the public by first establishing minimum criteria for operation, and then providing the public means of compensation in the case of an oversight by the institution. It also offers a measure of protection for the institution, providing a way for the restaurant to avoid being accused of health violations through producing records to a third party regulatory body.

Our proposal for an institution that intends to use a statistical database as justification for policy is to set up a similar kind of system where institutions that want to use the production of records as justification must meet some criteria first. The two criteria are simply Transparency and Verifiability.

I will go into more detail with transparency and verifiability using the CVS case here. These criterion form the basis of a new proposal of regulatory practices for institutions that wish to justify policy through statistical analysis. A society that successfully handles the difficulties implicit in institutional uses of statistical analysis must keep these systems transparent and publicly verifiable. If CVS expects to use their database to justify policy, then they must:

1. Be *transparent* in what metric they are using. Before implementing the policy, they must have some public announcement that they will be tagging products based on which products are reported stolen the most often.

2. Be *verifiable* in that the statistical database should either be public or sent as a record to some third party regulatory body to ensure that the enacted policy is actually in line with the stated policy.

For CVS, they would announce their new database system, and have the database in use be public access. They should also require that their CVS branches behave according to this broader policy in order to avoid the selective manager case, where a manager selects whatever policy is most detrimental to a certain class. The broad move here is to change how institutions view the deployment of statistics as part of policy. No longer should institutions see statistical information as convenient ways to deflect direct responsibility for their policies, they should instead treat the use of statistics in this way as a responsibility. What CVS is trying to get us to accept when they claim that their deference to their database freed them from wrong doing, is that we should treat statistical reports like temperature reports from a restaurant. Temperature reports from a restaurant, when the restaurant is accused of not meeting safety standards, can provide evidence for their guilt or innocence. But in order for that report to do that work, it must be produced and checked by a regulatory body. In addition, we demand that restaurants submit to regular health and safety checks that ensure these protections are in place. We can demand similar checks for institutions that wield large scale tools of statistical analysis.

3.6 Addressing the Problems

How do our two criteria address the concerns outlined in the beginning of this section?

Consider the masking problem first. In the masking case, we were concerned that the statistic cited by CVS were either constructed after the fact as a public relations move to limit the damage that the story was doing to their brand, or the policy based on the statistic functioned to help hide genuinely racist managers by offering them a choice as to what policy they prefer. As mentioned before, a manager might first check whether or not either protocol disadvantages black customers and base their decision on that determination. When accused of racism, the manager denies that they chose the protocol based on race and instead says that they were purely motivated by trying to keep the theft of objects at a minimum. With our two criterion, we address both issues. By stipulating that the institution publicly divulge exactly what they are using to inform their policy, and having the database publicly available, CVS cannot construct the justification after the fact, denying them the ability to hide behind statistical deference if they did not genuinely have such a system set up. The second problem of masking genuinely racist managers is avoided by having the policy be non-optional for managers of individual branches. Notably, this does not solve the issue, but rather kicks the issue up the chain of responsibility to whoever is deciding on the policy in the first place. In the original CVS case, we were worried that any branch manager could mask their racially motivated policy in addition to higher-ups in the CVS chain. In this case, at the very least we have limited the potential masking problem to just the people at CVS responsible for broad corporate policy. This is in line with the idea that institutions should be designed whenever possible to have the fewest possible avenues of exploitation.

Addressing concerns about statistics that are simple proxies for statistics on protected classes is more difficult. Just as we have a regulatory body responsible for setting the standards for operating restaurants, there should be a regulatory body responsible for determining which statistics are permissible for use as justifications. This is not as radical as it seems considering that our legal system already does this. For instance, several states prohibit employers from asking applicants how long it has been from their last job, partly to avoid discrimination against applicants that needed time off to recover from injuries. It is also commonly accepted the employers should not ask about marital status or whether or not the applicant intends to have children, again, partly to avoid discrimination against pregnant applicants. By extending these already granted powers to a more systematic regulatory body, we can attend to proxies on a case-by-case basis.

Habituation is harder to handle. It is partially addressed by having the institution make a public announcement of what exactly is being measured. Recall in our section on *Problems of Privacy and Implementation* that we considered the story of Eve and Franny.

> Eve and Franny are living under China's Sesame Credit. Eve is very knowledgable about the ins and outs of the system, and can explain in detail how the algorithm is generating the report, and what the report should look like given a brief description of the inputs. Franny on the other hand, knows very little about the mechanics of how the Social Credit System works, and like most other citizens, is only aware of very general rules that affect her score. She knows for instance, that her scores tend to rise when she pays back her debts on time, but there are fluctuations in her score that she cannot explain.

By having the institution make a public announcement, we can preempt the issue of habituation by giving the justification before the policy is observed. In addition, like in our Eve and Franny case, the transparency and verifiability criterion ensure that even if most of the population becomes habituated to this kind of tagging, we can have anyone act as a Eve to the system and check if the company is being honest with policy. Ideally, when we observe a potentially problematic policy, every customer could check if the policy falls out of some kind of established system, or if it constitutes a breach in anti-discrimination regulations. This may sound far fetched, but we can envision this working as a kind of fact check system.

It is not too hard to imagine habituating a population to check *snopes.com* whenever they come across a suspicious news story. If institutional policies were up to the same standard, it should be similarly feasible to habituate the population to check whatever system the regulatory body has in place to track an institutions policies based on statistics. Work would be needed to make habituation work for us rather than work for passivity. Habituating a population to actively investigate claims of statistical justification becomes possible with the transparency and verifiability criterion.

Our issues with normalizing behaviors can similarly be helped by our criterion. The worry that other institutions will hop on to a normalized behavior is avoided by having the system be verifiable for each institution, and by having a regulatory body. Just like in our restaurant case, the health and safety department comes by periodically to check that all the fire-extinguishers are in order and all the refrigerator temperatures are being properly recorded, we should expect our regulatory body check institutions to see if they are collecting the appropriate statistics honestly instead off copying them off of some larger institution. The issue of giving credence to groups like far-right nationalists will be addressed later. The last issue of normalization was concerned with insensitivity to changes in the statistic, but again, with the regulatory body in place, checking honest data collection would be as expected as checking for honest food safety measures.

Thinking broadly about the issue of sensitivity, this whole project can be seen as trying to establish conditions in which statistical evidence is sensitive. Recall earlier, when we were discussing Enoch Et Al.'s work, the fact that statistical evidence was insensitive to the particulars of the offense was a major strike against its use. With our conditions and the regulatory body in place, we can say that we have made the use of statistics somewhat sensitive to the particulars. To illustrate the parallel:

1. Alice's fingerprints on a gun, while being circumstantial evidence, is still sensitive to the particulars of a murder case because we can say if Alice did commit the crime, she would probably have her fingerprints on the gun, and if she did not, she probably would not have her fingerprints on the gun.

2. If we are in a discrimination case in our system, passing the checks provided by our regulatory body is sensitive to the particulars of the case because if the institution under examination was indeed being objectionably discriminatory it would be unlikely that the report provided would match precisely with the actions of the policy, with the mechanism of the policy being announced before its effect.

This system fits neatly under Enoch Et Al.'s as providing a method of dealing with statistical evidence that makes them sensitive to the particulars make it much easier to justify using them as compelling pieces of evidence in legal cases. No longer do we have the blue bus case where our statistic seemed divorced from the incident we were investigating.

Our last problem, with other accounts of racism, actually ties in with our previous worries about giving credence to groups like neo-nazi's. While our system can handle statistical justifications with background injustice fine, we have reason to pay special attention to socially pervasive forms of discrimination, i.e. forms of discrimination with background injustice. Two of the three presented accounts, the Responsibility Account and the Expressive Harm Account deal directly with background injustice, and we can use those, in conjunction with our system, to handle cases where the deployment of statistics could lend credence to extreme groups. In short, our system simply does not handle additional wrongs involving background injustice, and so we need to import accounts of discrimination that do in order to point out abuses of statistics where one group preys upon another group suffering from background injustice. In the section on other accounts of racism, I did some work trying to separate the kind of problem this system tackles and the kind of problem these other accounts of racism tackle, and here their separation serves as an advantage, as these other accounts apply where our system does not, and vice versa. Other accounts can handle the specific injustices that come with discrimination with a present background injustice, but my account can handle any claim that an institution is not truly motivating their policy with their claimed statistical analysis.

Chapter 4

Applications and Problems

In this chapter, we will take a closer look at what this system might look like in practice, and discuss issues of feasibility and what still needs to be developed.

4.1 The Regulatory Body

In the past few sections, it seems like I am relying heavily on a regulatory body analogous to health and safety organizations. I want to take a moment here to explore what this regulatory body should do.

I will fall back to the familiar example of restaurants and their refrigerator temperatures. Currently, restaurants are required to keep daily records of their refrigerator and freezer temperatures, and during any health and safety inspection, they must produce those records for review if asked.

Consider a refinement of this system. Instead of having to keep manual records, restaurants are instead required to keep a thermometer in each refrigerator, which records the temperature and sends it to the health and safety department daily. If we are to seriously consider incorporating more surveillance technology in our society, then this seems to be amongst the least contentious applications. In effect, the regulatory body sets standards that must be held, and checks them through inspection. The function of the health inspector and the surveillance mechanism is similar, to assist the regulatory body in determining if the institution is in line with regulations.

Setting up a regulatory body to track the use of databases is structurally similar to the health and safety case, but with some difficulties in deciding how to implement inspections and interpret the "thermometers". It seems clear that a thermometer in a refrigerator is the most sensible way to determine if the refrigerator is up to standard, but how do we decide if a policy and its associated database is up to standard? Let's go through the CVS case, paying attention to how the regulatory body will track policy.

CVS would like to use their database as justification for tagging products, and would like for their database to have protective force, i.e. they would like their deference to this database to be recognized as the sole determiner of policy regarding tagging procedures. In order for them to do so, they would first approach the regulatory body with their database and policy. To meet transparency, they would need to have available their policy and the database to at least the regulatory body, and to meet verifiability, the regulatory body must be able to, when asked, refer to the database and compare the results to the distribution of security tags. Analogous to the list of health and safety checks that a health and safety inspector might go through, this regulatory body might send inspectors to randomly selected CVS locations to help verify that the policy is properly in place.

We can question the feasibility of establishing a large regulatory body for the use of databases, but the regulatory body would not be necessary in most cases. CVS, instead of going through the regulatory body, might meet transparency and verifiability by posting their database publicly, and inviting anyone in any location to check that they are conforming to their security policy. I would imagine that most would take this route. This is fine, but the regulatory body has some other uses aside from ensuring our two criterion.

First, having a regulatory body can better handle databases that contain sensitive information. A company that wishes to establish policy on the hiring process, or an institution that has some policy that depends on where its members live, probably does not want to publish the relevant database online. We can instead have the institutions differ to a regulatory body. This step might appear to only push back the concern of privacy to concerns about the security of the regulatory body, but given that these demands only apply to institutions that attempt to implement policy, and wish for their deference to a database defend them from certain kinds of accusations, some vulnerability will be necessary.

Second, a regulatory body that sets standards for permissible uses of statistics allows us to have a sort of licensing procedure. Just as restaurants have a responsibility to the public that they must respect due to the vulnerability of the public to their food safety practices, institutions that create policies based on statistical evidence have a similar responsibility to the public. In the first case, the public is vulnerable to health concerns, and the government can intervene on their responsibility to public health and safety. In the second case, the public is vulnerable to discrimination, and the government can intervene on their responsibility to protect against discriminatory practices, already recognized by bodies such as the EEOC and acts such as the Civil Liberties Act of 1964. A restaurant needs licensing because it deals in food preparation, which improperly handled, can serve as a serious threat to public health, public health falling under the governments jurisdiction. Institutions that enact policies in deference to statistics needs licensing because it deals in the deployment of statistical evidence, which improperly handled, can serve as a serious threat to anti-discrimination laws, which fall under the governments jurisdiction.

4.2 Police Practices

We have taken a look at what I have constructed to be the easiest example possible, the CVS security tag case. Now, I will try to apply this system to a much more contentious case, Police practices.

Let us take a look at racial profiling. Suppose a police institution is accused of racial discrimination in how it determines its patrol routes. The police respond that they are merely patrolling in such a way that covers neighborhoods that are more likely to have crimes committed in them. Should we believe them, and in what cases must we accept that the police are indeed correct that their policy is actually pinned on the claimed statistic?

First, they must have been transparent, by announcing that they were planning on using this statistic before the policy was actually implemented in order to avoid the case that the police found the statistic after the fact. Second, they must also make the statistical evidence verifiable, possibly by making it public. If they have met this criteria, and we check the policy and find that it is and has been in line with their gathered statistical evidence, we should believe the police institution when they claim that the policy was made to lower crime rates, and not to discriminate in accordance with race. However, the police institution does not go off free. As mentioned earlier, racial profiling is a case with background injustice, in which the three accounts recounted earlier in the previous section are relevant. With a system that includes a regulatory body, we can also include the discussion of background injustice in the conversation of what kinds of statistical evidence the regulatory body should accept. When the regulatory body is deciding what kinds of statistics are appropriate for which kinds of policies, special attention can be made to protected classes.

In the discussion of police regulation, we seem to be comfortable with measures like police body cameras. Surveillance methods that would be inappropriate to demand on individuals seem permissible when surveilling institutions. We saw this kind of institution/individual asymmetry in the discussion of the restaurant and dinner party case, where the threat to public safety in a restaurant is considered to be more serious than that of a dinner party. As we consider measures like police body cameras, we should also consider measures not only to surveil the members of public institutions, but the policies of public institutions.

4.3 Employment

I will close this discussion by examining employability reports. We already have many restrictions on what employers may consider when thinking about who to hire. While generated hiring reports are already being used by large companies, it seems like they are much more hesitant to automate their hiring process like one might be tempted to automate the tagging of items at a CVS. However, even procedures that are designed to filter out applicants can be problematic. Additionally, we have at least one example of an automated employment process that was mentioned earlier. The pseudo-taxi company Uber automatically fires drivers when their approval scores dip below a certain threshold.

Consider our system approaching something like “days since last job” as a proposed statistic to use in hiring procedures. While this may seem innocent, this statistic has been accused as a proxy for discriminating against recently pregnant women and the chronically ill, who may need extended time off from work to attend to other responsibilities. Currently, many states have responded by preventing employer from posting “unemployed need not apply” conditions in their applications. In this system, the regulatory body should catch this attempted proxy and deny that the statistic is appropriate for use in policy. The job of a CEO determined to discriminate against pregnant women through deference to statistics becomes much harder. They need to find some statistic which will pass the scrutiny of the regulatory body that will reliably discriminate against pregnant women and hold up over time as the statistic is published publicly and continually verified. The statistic should also be a believable measure for competency in the job at hand. While finding such a statistic is not impossible, it is, at the very least extremely difficult.

Chapter 5

Conclusions

We set out to investigate the role of statistics in justifications for institutional policy. First, we took a look at an existing precedent, market share liability, to see if it would generalize to help with the use of statistics in justifying institutional policies. Although the principle did generalize, market share liability is not suited for handling statistical evidence in the context of institutional policies. We identified some unique problems and potential abuses that institutions that defer to statistics might commit, and argued that the use of large scale databases to form policy needs regulation if it is to have force in the court room. Specifically, if we are to believe an institutions claim that a policy relies on some deference to a statistic, the policy and the statistic should be transparent, starting from before the implementation of the policy, and verifiable, for however long the policy is in place. The broader point I have tried to make is that there is something deeply wrong with how easily we accept deference to a statistic to be a morally neutral action. We should give large institutions a *very hard time* when they attempt to excuse their actions through deferring to a statistic. It is easy to hide embarrassing, damaging, or irresponsible actions when all you have to do to appear free from sin is produce a believable statistic that supports the action.

Our capacity to collect, analyze, and deploy statistical evidence is growing quickly, and with it, grows our capacity to keep track of how institutions are using statistical evidence. I can imagine a few years down the line, something like CVS's claimed database will pop up, and when it inevitably produces apparently discriminatory results, we should be prepared to buckle down and investigate what is going on in these, sometimes murky, systems. I believe that it is a near inevitability that systems that track and collect a large amount of statistical evidence will start to include analysis features that are meant to inform policy. Crystal Commerce, an inventory tracking system, is only a few lines of code from offering a sorted list of most often stolen items. Salesforce, a customer relationship tracking program, is similarly only a few lines of code from suggesting which classes of people donate the most and to what cause. We should work hard to set the expectation that these large scale systems stay transparent and verifiable, for the analytic power that these systems provide, combined with the institutional power given to policy makers, make for a dangerous combination.

Bibliography

[1] Bou-Habib, Paul. ”Racial Profiling and Background Injustice.” *The Journal of Ethics*. 15, no. 1/2 (2011): 33-46. http://www.jstor.org/stable/41486899.

[2] Cory Doctorow. *Little Brother*. (New York: Tom Doherty Associates LLC), 47.

[3] Court of Appeal[2000]: Clark, R v [2000] EWCA Crim.

[4] “CVS Hair Care Products,” last modified June 22, 2014, www.snopes.com/racial/business/cvs.asp

[5] David Enoch, Levi Spectre, and Talia Fisher, “Statistical Evidence, Sensitivity, and the Legal Value of Knowledge.” *Philosophy and Public Affairs*. 40 (2012).

[6] Exec. Order. No. 10925, 26 FR 1977 (March 8, 1961), https://www.archives.gov/federal-register/executive-orders/1961-kennedy.html.

[7] Frank Pasquale. “Judge, Jury and Executioner: the unaccountable algorithm.” *Aeon*. August 18, 2015.

[8] Federal Tort Claims Act, 28 U.S.C. Pt.VI Ch.171 (1946)

[9] Judith Thomson, “Liability and Individualized Evidence.” *Law & Contemporary Problems*. 49 (1986).

[10] Richard Schmalbeck, “The Trouble with Statistical Evidence.” *Law & Contemporary Problems*. 49 (1986).

[11] Robert Howell, “Google Morals, Virtue, and the Asymmetry of Deference.” *Noûs*. 48:3 (2014): 389-415.

[12] Sindell v. Abott Laboratories, 449 U.S. 912 (1980).

[13] *State Council Notice concerning Issuance of the Planning Outline for the Construction of a Social Credit System*. (2014-2020), translated by Roger Creemers.

[14] Wyeth v. Levine, 555 U.S. 555 (2009).