

The Interaction between Social Science and Judicial Reasoning in the Context of Employment Discrimination Cases

Xiaorui (Sherry) Xia

Legal Studies Honors Thesis Draft

University of California, Berkeley

Spring 2014

Under the supervision of:

Catherine Albiston

“For the rational study of the law the black letter man may be the man of the present, but the man of the future is the man of statistics and the master of economics.”¹

-Oliver Wendell Holmes

¹ "The Path of the Law" 10 *Harvard Law Review* 457 (1897)

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my thesis advisor, Catherine Albiston, for her continued support and guidance throughout every stage of this project. I would also like to thank Professor Michael Musheno and Christina Stevens-Carbone for their generous feedback and encouragement.

ABSTRACT

The use of social science evidence in court arose out of an expectation that social science evidence can offer new insights to aid dispute resolution in court. Yet fundamental differences between scientific reasoning and legal reasoning have led to serious challenges in incorporating such evidence in legal decision making. In particular, these differences include the discrepancy between legitimacy criteria in law and social science, as well as controversies that arise from general-to-specific causation.

In light of the discrepancies between scientific and legal reasoning and the impact of the Daubert decision, I explore the treatment of social science expert testimonies in the context of employment discrimination cases that violate Title VII of the Civil Rights of 1964. I hypothesize that challenges between scientific and legal reasoning make judges reluctant to admit social science evidence in the courtroom.

Drawing upon employment discrimination cases from the Federal District Court over the period of 1993 to 2014, I look at the admissibility of social science expert testimonies as well as the use of statistical evidence in court. I use this analysis to identify judges' attitudes toward the scientific evidence and how social science evidence is used alongside legal reasoning. I find that despite of distrust of the legal system in the use of statistical evidence in the courts, judges have become more comfortable with the use of statistical evidence in court.

Table of Contents

I. Introduction

II. Admissibility and the Tension between Science and Legal Reasoning

- Social Science and Causal Inference
- Social Science and Mutability
- The Role of Judges in Interpreting Social Science
- Judges and their Attitudes towards Social Science
- Judges' Ability to Apply the Daubert Standard
- Debate in Scholarship

III. Methodology

IV. Analysis

V. Limitations

VI. Conclusion and Further Questions

VII. Bibliography

I. Introduction

Social science evidence found its way into the courtroom as a tool to account for and predict human behavior. Jurisprudential origins of social science in American law can be traced to the emergence of empirical inquiry during the rise of Legal Realism in the mid-20th century. Leading the Legal Realist movement was a cry against the Legal Formalism of the 1870s and 1920s that perceived law as a series of rules and fixed axioms. The rise of Legal Realism and thinkers such as Holmes, Brandeis, and Pound, criticized classical jurisprudence and led to the motivation to study law in relation to society itself. As Oliver Wendell Holmes famously stated “[t]he life of the law has not been logic: it has been experience” (1881, p.1), the rise of the Legal Realist movement led to the study of the behavior of society at large in conjunction to law.

Social science evidence was first introduced in the courts during the Progressive Era to expose the disagreeable working conditions of urban workers. Social science evidence found its way to the US Supreme Court in *Muller v. Oregon (1907)* where the constitutionality of an Oregon statute that limited the working hours of women was challenged. Social science evidence that was presented in the Brandeis brief consisted of a compilation of social science facts including testimonies by social scientists, medics, and male workers that long working hours had a significant negative effect on the health, safety, morals, and general welfare of women. An excerpt from the brief reads:

The experience of manufacturing countries has illustrated the evil effect of overwork upon the general welfare. Deterioration of any large portion of the population inevitably lowers the entire community physically, mentally, and morally. When the health of women has been injured by long hours, not only is the working efficiency of the community impaired, but the deterioration is handed own to succeeding generations. Infant mortality rises, while the children of married working-women, who survive, are

injured by inevitable neglect. The overwork of future mothers thus directly attaches the welfare of the nation.²

Despite of legal precedents such as *Lochner v. New York (1905)* that strongly supported striking down the Oregon statute as being unconstitutional, the Court relied on the Brandeis brief to rule in favor of Oregon's statute, justifying it as necessary for the protection of the health and welfare of women. The brief pioneered the introduction of social science evidence into law and paved the way for more social science evidence to be introduced into the court system.³ The use of empirical analysis that grew out of *Muller* existed as a methodological alternative that offered a fresh perspective to the doctrinal analysis of law by looking beyond the law books and towards how society operates as a whole. Social science and the social facts can serve to answer to legal problems and guide judges in their decision making.

Optimism surrounding its usage can be seen from Oliver Wendell Holmes' prediction that while "the rational study of the law the black letter man may be the man of the present, but the man of the future is the man of statistics and the master of economics".⁴ Like Holmes, legal scholars of the time saw social science evidence as an indispensable tool that can revolutionize the way things had traditionally been done in court.

Today, a more sophisticated understanding of the methods of science has led to a more critical approach towards the use of social science evidence. In particular, changes in social science methodology have led to a scrutiny of the usage of such evidence in the past. Brandeis' brief in *Muller v. Oregon* for example would not be acceptable today as admissible social science

² Brief for the defendant in error *Muller v. Oregon*.

³ Bernstein, David E. *Rehabilitating Lochner: Defending Individual Rights against Progressive Reform*. Chapter 4. Chicago: University of Chicago Press. (2011): 195.

⁴"The Path of the Law" 10 *Harvard Law Review* 457 (1897)

evidence. Some of the challenges that are raised against the use of social science evidence arise from inherent differences between science and law and relate to discrepancy between legitimacy criteria in law and social science as well as controversies that arise from general-to-specific causation. For example, serious challenges have been posed against the use of the *Clark Doll Experiment* (1939) in *Brown v. Board Education* (1954). Critics find fault in the methodology of the experiment as well as the conclusion that the experiment reached⁵. Similar challenges to the validity of social science evidence in other cases pose serious threats to the legitimacy of legal precedents that rely on social science evidence.

To address these differences, scholars have made major efforts to frame social science in the context of law. Notably, Monahan and Walker have dedicated much of their work to delineating the role of social science in law. Despite of the efforts the made in the field, the incorporation of social science evidence into the legal framework in practice still face many of the preceding challenges.

For the purposes of this study, I would like to explore the challenges of incorporating social science evidence in law and aim shed light on judges' attitudes toward the use of social science evidence in court. I explore the treatment of social science expert testimonies in employment discrimination cases that violate Title VII of the Civil Rights of 1964. I look at the admissibility of social science expert testimonies as well as the use of statistical evidence in employment discrimination cases from the Federal District Court over the period of 1993 to 2014. I analyze judges' attitudes toward the scientific evidence and how social science evidence is used alongside legal reasoning in these cases to shed light on how challenges between scientific and legal reasoning

⁵ Duizend, Richard, and Saks Michael. *The Use of Scientific Evidence in Litigation*. Williamsburg: National Center for State Courts, 1983. Print.

operate in these contexts. I hypothesize that challenges between scientific and legal reasoning make judges reluctant to admit social science evidence in the courtroom.

II. Admissibility and the Tension between Science and Legal Reasoning

In the following sections I outline theories and principles behind the challenges of using social science evidence in court. I begin by laying out the theoretical foundation for the inherent tensions that exist between science and legal reasoning. Then I go on to show how Judges' roles as gatekeepers to social science evidence and their attitudes towards social science evidence matters in determining the role of social science evidence in law. Finally, I show that despite of contributions in scholarship to delineate the role of social science in the court, there are still areas of contention that fails to be addressed.

Social Science and Causal Inference

Science and law have different understandings of causal inference. While scientists usually study variables at the population level, the Court aims to determine the particulars in a case. This idea is best summarized as evidentiary incommensurability which describes the difficulty of using scientific evidence in court due to the inherent differences between science and litigation.⁶ When scientific evidence is used in court, there is a question of how to translate generalizations drawn from the population to apply to the specific instances in question, especially when an expert testifies about general research that did not involve the parties in the case and then proposes to apply such research to the specific case at hand. In *Wal-Mart v. Dukes* (2011), the court was not satisfied with the social science evidence provided by expert witness Bielby's testimony⁷ and argued that statistical evidence as presented could not demonstrate that individual members of the class action experienced similar discrimination. This, however, is

⁶ Faigman, David. *Evidentiary Incommensurability: A Preliminary Exploration of the Problem of Reasoning from General Scientific Data to Individualized Legal Decision-Making*, 75 BROOK. L. REV. 1115, 2010.

⁷ Expert Report of William T. Bielby, *Betty Dukes, et al. v. Wal-Mart Stores, Inc.* Feb. 3, 2003

something that statistical evidence simply cannot provide. While it is unclear whether the Court made this decision due to its lack of understanding of statistics or used this as a strategic move to defeat the claim, whatever the intention of the Court was, the ultimate effect of the Wal-Mart decision is to set a high bar for the admissibility of social science expert testimony in court.

Likewise, while social science evidence can prove correlation, rarely can it prove causation. In *Brown v. Board of Education*, social science evidence was used to prove that segregation in schools caused Black children intellectual and psychological harm.⁸ Critics argue that even if harm had been demonstrated by the test, it was not clear how school segregation apart from other forms of segregation and discrimination in society could have been identified as the cause of the harm. As shown by these examples, the courts face considerable challenge in applying social science evidence in legal reasoning, for while science deals in probabilities, the court deals with the determination of outcomes in individual cases.

Social Science and Mutability

Social science and legal reasoning also differ in their relation to stability. While science is mutable, the law is not. A legal system based on precedent is made to be stable and consistent. In science, theories are made to fall away as more sophisticated theories replace them. As knowledge accumulates, new theories emerge to replace the older theories that accounted for the earlier knowledge base. In science, progress creates legitimacy, whereas in law, precedent creates legitimacy.⁹ As science constantly experiences change and innovation, embedding certain notions of scientific principles into the law will hinder the continuity of the law as theories embedded in law are resistant to change even when they are wrong.

⁸ Duizend, Richard, and Saks Michael. *The Use of Scientific Evidence in Litigation*. Williamsburg: National Center for State Courts, 1983. Print.

⁹ Woodward, James, "Scientific Explanation", *The Stanford Encyclopedia of Philosophy* (Winter 2011 Edition), Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/win2011/entries/scientific-explanation/>>.

The Role of Judges in Interpreting Social Science

Traditionally, trial court judges exercised some control over the decision making process of the jury. From 1923 to 1993 the Rule 702 of the Federal Rules of Evidence was governed by the Frye Standard which established the general acceptance test as a basis for assessing the admissibility of scientific expert testimony.¹⁰ The general acceptance test aimed to assess the reliability of expert scientific testimony by looking to see if the methodology or technique used to obtain the evidence is generally accepted as reliable in the relevant scientific community. The standard however was replaced by *Daubert v. Merrell Dow Pharmaceuticals, Inc.* (1993), which set a higher standard of admissibility. The Daubert standard consists of a two-pronged test based on reliability and relevance of the evidence. The judge's role was to determine whether 'an expert's testimony both rests on a reliable foundation and is relevant to the task at hand,' not on the expert's conclusions but on the 'principles and methodology' used.¹¹ As the gatekeepers, District court judges have the responsibility of determining what type of scientific expert testimony is appropriate.

To determine whether an expert's reasoning and methodology is reliable, courts can look to a diverse range of factors including whether the theory or technique on which the expert relies has been or could be tested or whether the theory or technique has been subjected to peer review and publication.¹² The implications of *Daubert* are huge as it implies that judges should not simply defer to members of a scientific discipline when deciding whether to admit expert testimony, but that they should be the ones to separate good from bad science. This dramatically changed a judge's responsibility from having to merely assess the general acceptance of a study

¹⁰ *Frye v. United States* (1923)

¹¹ Edward K. Cheng, David L. Faigman, Michael J. Saks, and Joseph Sanders. *Modern Scientific Evidence: The Law and Science of Expert Testimony*. Eagan, MN: Thomson/West, 2006

¹² *Daubert*, 509 U.S. at 593-94, 113 S.Ct. 2786;

as established in *Frye v. US (1923)* to having to evaluate the research methods and the analyses that lay behind expert opinion. Judges therefore have a lot of responsibility in determining what social science evidence gets admitted under this new role.

Studies on the impact of the *Daubert* decision show that in practice Judges are now more rigorous in determining the admissibility of expert testimony and applying the *Daubert* standard. A new report from RAND studying the impact of the *Daubert* standard on Federal court decisions examined the effects of this ruling on the behavior of federal judges, plaintiffs, and defendants.¹³ They found that judges have been taking their new role seriously and applying stricter standards to determine what sort of expert evidence is admissible. The report found that lack of general acceptance was as much a barrier to admission after *Daubert* as it was before. What's more, they found that general acceptance was no longer sufficient for evidence to be found reliable. After *Daubert*, judges began to find evidence unreliable even though it was generally accepted. Judges were not only more rigorous about assessing reliability of the evidence after *Daubert*, they were also more rigorous about assessing relevance and expert qualifications. In fact, according to RAND, judges began to examine all dimensions of the evidence more closely, leading to an overall more careful inquiry even on the traditional factors pre-*Daubert*.

Judges and their Attitudes toward Social Science

Not surprisingly, according to a 2002 RAND study, the percentage of expert testimony by scientists that was excluded from the courtroom significantly rose after the *Daubert* decision. While this study was conducted over court cases involving all legal areas, and not just employment discrimination, it is probable that this phenomenon also carries over to employment

¹³ Research And Development Global Research and Policy Think Tank, *Changes in the Standards for Admitting Expert Evidence in Federal Civil Cases Since the Daubert Decision* by Lloyd Dixon and Brian Gill

discrimination cases with regard to the admissibility of social science expert testimony. Legal scholarship indicates that Judges tend to be skeptical in admitting social science evidence in the court room.¹⁴ If social facts are needed to decide a particular case, judges tend to take judicial notice of ‘common knowledge’ or to rely on personal experience, rather than to use the findings of scientific investigations. Krieger & Fiske suggest that judges would rather use common sense to influence their decision making than to look at empirical notions of the world as it is. In this way, judges become “intuitive psychologist[s] behind the bench,” using common sense theories about what people do to influence their decision making, even though they are usually wrong.¹⁵ The judicial proclivity to make observations or behavioral predictions without empirical support drives judges to reject or overlook relevant social science evidence.

Judges’ Ability to Apply the Daubert Standard

Chief Justice Rehnquist with whom Justice Steven joins, writes in his concurring in part and dissenting in part opinion in *Daubert* expressed serious concern with the majority’s confidence in judges’ abilities to adequately judge the admissibility scientific evidence:

The Court speaks of its confidence that federal judges can make a ‘preliminary assessment of whether the reasoning or methodology underlying the testimony is scientifically valid and of whether that reasoning or methodology properly can be applied to the facts in is. I do not doubt that Rule 702 confides to the judge some gatekeeping responsibility in deciding questions of the admissibility of proffered expert testimony. But I do not think it imposes on them either the obligation or the authority to become amateur scientists in order to perform that role.

¹⁴ In E. Borgida & S.T. Fiske (eds.), *Beyond Common sense: Psychological Science in the Courtroom*. (pp. 127-156). Oxford: Blackwell Publishing.

¹⁵ Krieger & Fiske *Behavioral realism in employment discrimination law: Implicit bias and Disparate Treatment*. California Law Review, 94(4), 997-1062

The difficulty in the application of this standard has produced real concern as to whether or not federal court judges have the proper training to ensure scientific quality of expert scientific testimony. Some commentators believe that *Daubert* caused judges to become amateur scientists, many lacking the scientific literacy to effectively fulfill their role as gatekeeper of scientific evidence. Although “science for judges” forums have emerged in the wake of *Daubert* in order to educate judges in a variety of scientific fields, many are still skeptical about the usefulness of the *Daubert* standard in discerning valid science. The responsibility to assess scientific relevance has shifted from highly trained expert witnesses to judges deficient in science education. The "Daubert" ruling furthermore admits the possible introduction of non-peer reviewed data and conclusions. This increasingly shifts the burden of scientific judgment onto judges who have not had an education which would enable them to properly evaluate such data.

Existing studies raise concerns about judges’ ability to perform the Daubert requirement. A study found that only 4-5% of judges accurately understand the Daubert criteria of falsifiability and error rates (Gatowski et al 2001)¹⁶. In addition, another study that sought to test whether judges are sensitive to variations in the methodological quality of scientific evidence found that methodological quality influenced judges’ ratings of scientific quality and admissibility decisions very little to not at all (Kovera & McAuliff, 2000)¹⁷. These studies and more challenge the assumption that judges are able to operationalize *Daubert*’s scientific concepts and appropriately use them in their decision making.

¹⁶ Gatowski, Sophia I.; Dobbin, Shirley A.; Richardson, James T.; Ginsburg, Gerald P.; Merlino, Mara L.; Dahir, Veronica *Asking the gatekeepers: A national survey of judges on judging expert evidence in a post-Daubert world*. Law and Human Behavior, Vol 25(5), Oct 2001, 433-458

¹⁷ Kovera MB, McAuliff BD *The effects of peer review and evidence quality on judge evaluations of psychological science: are judges effective gatekeepers?*. Journal of Applied Psychology, 85(4), Aug 2000, 574-86.

There is a question however of whether these statistics properly applies to social science evidence where there is less rigorous requirement for a background in hard sciences, which judges are not usually equipped with. The current study focuses on social science as opposed other types of scientific evidence that are introduced in court and aims to see whether determining admissibility of social science evidence plays to judges' favor. Similarly it is possible that judges might hold preconceived skepticism towards social science evidence as it is not as established as other types of scientific evidence.

Debate in Scholarship

Scholars in the social science community often propose their own views about how the court should address challenges in the use of such evidence. While these insights are often helpful, disagreements in scholarship as to what is the appropriate action for the court in addressing certain types of social science evidence can confuse judges more than it can help them in determining the proper course of action. In areas where scholarship is in disagreement, the court faces the heightened difficulty of deciding what to do with the social science evidence.

One such debate in social science evidence concerns the use of social framework evidence in court. Social framework evidence is used to explain relevant, general, social and behavioral science research, and illustrate its applicability to the particular context of a case. It provides fact finders with a general empirical background or context within which to determine specific facts at issue in a case. The paradigmatic linkage question is presented when an expert testifies about general research that did not directly involve the parties in the case before the court. Controversies have arisen over whether the use of this type of framework evidence is admissible in court.

In “Contextual Evidence of Gender Discrimination,” John Monahan, Laurens Walker, and Gregory Mitchell argue that courts should never let social scientists link general social science findings to an employer's specific workplace policies, unless the expert has conducted his or her own empirical research in that particular workplace. They argue that admissibility of this kind of expert testimony would be contrary to basic principles regarding the admissibility of expert testimony. The opposing camp argues that such evidence should apply to social framework expert testimony, and that the flexible and permissive standards of the Federal Rules of Evidence allow framework testimony to be offered by a qualified expert in employment class action suits. The debate over the admissibility of social framework evidence adds to the inner struggle that judges’ have concerning the use of social framework evidence. So far, no clear consensus has been established in scholarship as to what is the best measure for judges to take in the use of social framework evidence.¹⁸

III. Methodology

I want to look at judicial reasoning in the context of social science evidence by looking at the admission or exclusion of expert scientific testimony in employment discrimination cases.

Choice of Judicial Opinions

To get an idea of the judicial reasoning of judges I use data coded from reported¹⁹ judicial opinions in federal district court cases over a 22 year period from 1993 to 2014. Reported opinions cannot fully capture the actual mental processes of the judges who wrote them but the decisions judges make with regard to the exclusion or admission of evidence and their

¹⁸ Monahan, J., Walker, L., & Mitchell, G. (2009). *The Limits of Social Framework Evidence. Law, Probability, and Risk*, 8(4), 323-351.

¹⁹ Not all cases are reported and when a court renders an opinion, it may order that the opinion be published in official case reports, it may make the opinion generally available for public distribution but specify that it be considered legally unpublished and therefore not cited as precedent, or it may simply file the opinion, in which case it generally does not appear either in official reporters or in the online database. The Westlaw federal database is based on extensive efforts to include all published and (legally) unpublished cases (Edelman et al. 1999)

justification for such evidence serve as a good platform to gauge their thoughts.²⁰ It is through these opinions that judges make their legal arguments in relation to the facts that are presented to them. Therefore, I look at judicial opinions as a reflection of judges' perception of the case as a whole and hope to get their understanding of scientific testimony.

Choice of District Court Cases

I excluded the study of US Supreme Court cases and Appellate Court Cases and focus instead on District Court cases. I excluded cases from the US Supreme Court because of their relatively small number as Daubert admissibility of evidence does not usually factor into Supreme Court decisions. If they do, these cases would best be treated as independent variables in order to examine their effect on judicial reasoning in the circuit and district courts. The federal Appellate Court²¹ cases on the other hand do not deal with Daubert motions explicitly and rarely discusses it in much detail. It is in the federal District Courts where Daubert comes into play.

So, another reason to focus on district courts is because Daubert decisions are decisions about whether or not to admit evidence at the trial level. Accordingly, most of those decisions will be made in trial courts. Also, when a party appeals a decision about admissibility of evidence, the appellate court generally gives significant deference to the trial court's decision. As a result, very few decisions about admissibility end up being appealed because generally the appeal will fail, although not always. If the error is not "harmless", e.g., it likely affected the outcome of the case, it could be reversed. But that's a tough standard to meet.

Case Selection

²⁰ Published journal articles also address some of this and I do examine some of those articles in my literature review, but they do not always depict how judges deal with these issues in action. The literature on this matter covers many of the ideas that judges have think about or even try to apply in practice, but it is not always what judges do. Opinions capture judges attitude and reasoning in action in the face of complex litigation.

²¹ I began by looking at appellate court decisions only to find that they do not deal with *Daubert* as much as I would like. I began by looking appellate decisions, but realized that that appellate court decisions can't be relied upon to generalizable to all cases.

I used the Westlaw²² database to select all reported federal District Court cases decided by the U.S. District Court between 1994 and 2014.²³ I also chose to frame my Westlaw search to the period from 1994 to 2014 because *Daubert v. Merrell Dow Pharmaceuticals, Inc.* was decided in 1993. This search yielded a total of 468 district court opinions.

My search term specified that cases be brought under Title VII of the Civil Rights Act of 1964. I chose not to look at other forms of discrimination claims brought under statutes such as the Americans with Disabilities Act of 1990 (ADA) as well as Age Discrimination Employment Act of 1967 (ADEA). I chose to only look at claims brought under Title VII of the Civil Rights Act of 1964 to frame my questions so as to not make my question too broad and also to ensure that these cases all involved comparable legal questions. Although I chose to only look at Title VII discrimination cases and excluded other discrimination cases, the majority²⁴ of the forms of discrimination are Title VII discrimination cases. Such claims include discrimination based on race, color, religion, sex, as well as sexual harassment. Many cases that were generated by the search raised claims under multiple statutes. I included any cases that involved any of Title VII claims.

My search term also specified that I would be looking for cases that involve the application of the Daubert standard. To do this without being over exclusive I included “Daubert” short for *Daubert v. Merrell Dow Pharmaceuticals, Inc.* and “expert testimony” in my search term.

²² Westlaw is one of two comprehensive databases of judicial decisions. I chose Westlaw over Lexis primarily because it was easier to import data into SAS and to order the cases chronologically. I also had free access to Westlaw vs. Lexis.

²³ My search terms include "expert testimony" & "title vii" & "dow" and I set the following filters: Federal District Court Cases, Further filtered by: Reported and Employment & Labor Cases, Generated a total of 745 cases, Look at district courts, talk about how they weigh them.

²⁴ EEOC.gov lists the different types of discrimination to include age, disability, equal pay/compensation, genetic information, harassment, national origin, pregnancy, race/color, and religion, retaliation, sex and sexual harassment. Title VII covers the majority of these and by far the most common ones.

Qualifying Features

As I intentionally used a broad search term in order to include all possible opinions, I had to go back to qualify the opinions by checking to see if the cases met my criteria. My qualifying criteria included the following (1) the case cites *Daubert v. Merrell Dow Pharmaceuticals, Inc.*(1993) , (2) the case deals with one of the employment discrimination claims brought under Title VII of the Civil Rights Act of 1964, and (3) cases that had been overruled or superseded, and therefore is no longer good legal authority I found applying these qualifying criteria that many opinions initially selected by my search term were not actually employment discrimination cases and that many more did not cite *Daubert v. Merrell Dow Pharmaceuticals, Inc.* (1993), but merely had a plaintiff or defendant by the name of Daubert. To find out if the cases included a discrimination claim that fell under Title VII discrimination claims (Civil Rights Act of 1964) I coded for the discrimination claims that each case made. I found that many of them included ADA, pregnancy, and other forms of discrimination that did not fall under Title VII. I excluded those as well.

I also excluded cases that had any negative treatment. Westlaw traditionally provides keys that indicate if any cases had negative history. I excluded cases that are flagged red but not the cases that are flagged yellow in Westlaw. A Westlaw KeyCite that was flagged red²⁵ means that the case has been overruled or superseded, and therefore is no longer good legal authority. A Westlaw KeyCite flagged yellow²⁶ means that the case or administrative decision has had some negative history, meaning other courts disagreed with it, but hasn't been reversed or overruled. I looked at the cases that were flagged yellow on an individual basis as these cases are probably

²⁵ A red flag in cases warns that the case is no longer good law for at least one of the points it contains. In statutes, this flag indicates that a section has been amended or repealed by a session law or that it has been declared unconstitutional or preempted.

²⁶ In cases and administrative decisions, a yellow flag warns that the case or administrative decision has some negative history but hasn't been reversed or overruled.

the cases in which the courts are getting it wrong, and so would be of the most interest to this project. Yellow flagged cases are those where there is confusion and controversy about how to apply the standard, and that's what is interesting to my project.

If decisions did not meet the above three criteria, they were excluded. This rigorous exclusion process eventually led me to a total of 49 cases.²⁷ As this qualifying criteria led to an extremely small number of cases, I did not use sampling techniques.²⁸

General Data

My coding scheme includes data only from the opinion level. At the opinion level, I collected general data on the case name, date of decision, the name of the judge, whether *Daubert v. Merrell Dow Pharmaceuticals, Inc.* was cited in the case, and a variety of other factors.²⁹

Plaintiff Data

I also coded for some information on the plaintiffs. Specifically, I coded for the sex and race of the defendants in relation to the claims they brought to court. I did not always code for these criteria for plaintiffs who did not bring a corresponding employment discrimination claim to court. For example, I did not code for the race of a person who did not make a race discrimination claim. This was because plaintiff's race was not always available in the cases involving claims of gender discrimination cases and vice versa. I also coded for whether the plaintiff was a public interest organization or government organization or neither. I chose to focus more on the plaintiffs because most Daubert motions are usually brought by the defendants

²⁷ I excluded 346 cases based on the first criteria, 28 cases based on the second criteria, and 43 cases based on the third criteria.

²⁸ Originally I sampled from the original search result of 468 cases. I ordered all opinions in chronological order and randomly selected a sample 211 district court opinions. However as I tried to qualify the sample, I found that very few cases actually qualified. This led me to drop my sample and to expand my sample size. Eventually, I had to use all the cases in the population of cases that fell under my search term.

²⁹ Other factors include the district court in which the decision was made, the statutory claims involved in the case, and whether an amicus brief was filed.

against the plaintiffs. In addition, plaintiffs are the ones who must bear the burden of proof when making a claim.

Legal Theory

For each opinion, I also coded to see if any legal theories were mentioned in the opinion. I took note to see if the discrimination claims were made in the context of disparate impact or disparate treatment theories respectively. I coded for whether disparate impact or disparate treatment theories were made by the plaintiffs. This was of an interest, because much statistical evidence is related to disparate impact and disparate treatment claims. I did not code for who won on the claims, but took note of how the court discussed the adequacy of the each claim based on these legal theories when it was related to expert testimony or the use of statistical evidence.

Expert Testimony

As I want to look at judicial reasoning with regard to the admissibility of expert testimony, I code for whether expert testimony was admitted, partially admitted, or were excluded. While the majority of cases involved determining the admissibility of one expert witness, others involved the admissibility of multiple expert testimonies. The admissibility of each expert's testimony was coded for each case. If the case expert's testimony was struck down or partially admitted, I took note of the criteria judges found the testimonies failed to meet. Such criteria include relevance, reliability, ability to assist the jury, and etc. I only coded for the admissibility of expert testimonies for testimonies that were introduced by the plaintiffs, but not for the defendants. In most employment discrimination cases it is the plaintiffs who bring in experts, and not the defendants. However the defendant occasionally brings in their own experts to challenge the expert testimonies brought in by the plaintiff. I coded for whether the defendant

brought in their own experts to testify in the case and looked to see whether that impacted the admissibility of the plaintiff's expert testimonies.

Statistical Evidence

Aside from the admissibility of expert testimony, I also coded for the existence of significant discussion of statistical evidence in the opinions as I want to understand how judges deal with such evidence. I coded this based on several criteria: (1) whether statistical evidence was explicitly mentioned in Westlaw Headnotes,³⁰ (2) that there is at least one to two paragraphs discussion solely on one type of statistical evidence in the opinion, and (3) if a word search of the word "statistic" brings up no search results, I categorized the case as not having any significant discussion of statistics.

One risk of this strategy is that it might be over inclusive. I was careful about reading the relevant passages on the discussion of statistical evidence and followed the judge's reasoning there. Nonetheless, it is difficult to determine objectively what counts as having a significant discussion of statistical evidence.

In addition to coding for the presence of significant statistical discussion in the opinions, I also coded for the type of statistical evidence that the discussion was based on. For example, whether it was based on expert testimony regarding the expert's methodology or it was on independent statistics brought by an outside source.

Coding for Themes

In addition to coding for most of the quantitative data I also code for qualitative data by capturing themes that appear in the opinions. I conducted a detailed reading and analysis of all 49 cases in my population and kept track of common themes that appeared across the cases. To

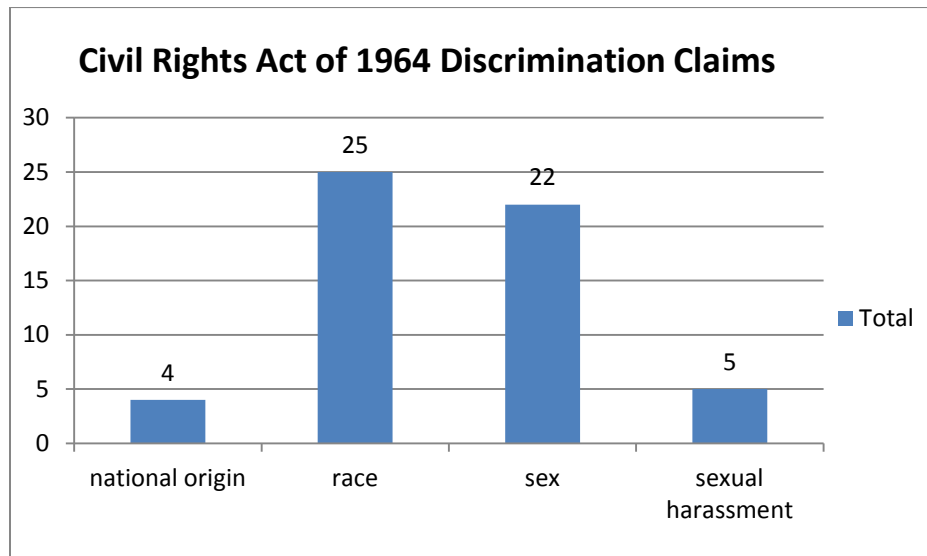
³⁰ Westlaw Headnotes is a service provided by Westlaw which outlines the major judicial points mentioned in an opinion. If statistical evidence is mentioned in the headnote, there was usually a significant discussion of statistical evidence in the body of the opinion.

ensure consistency in identifying themes I looked at one case at a time and went go back and forth between cases to capture popular themes and motifs.

IV. Analysis

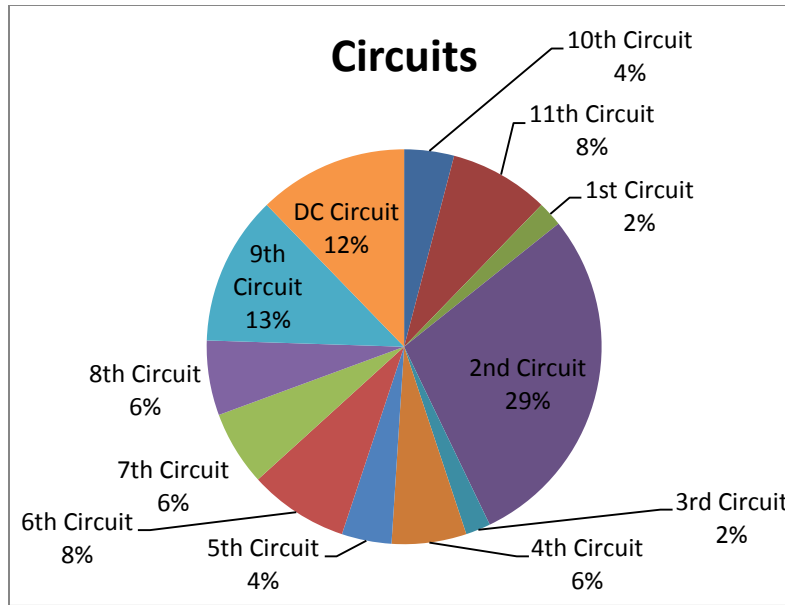
Discrimination Claims

Of the 49 cases in my population, the majority of the cases concern some sort of race or sex discrimination claim. There are 25 claims of discrimination based on race and 22 claims of discrimination based sex.³¹



Circuits

³¹ I coded for discrimination claims based on the claims that were made



One thing that stood out in my analysis was that a high percentage of cases come from the 2nd Circuit. In fact, 14 cases, making up 29% of the total cases in my population, come from the 2nd Circuit. Of these, 12 cases come from the New York district courts and 2 cases come from the Connecticut district courts. It is also interesting that 10 out of the total 14 cases that come from the 2nd Circuit come from the Southern District Court of New York.

The high percentage of cases seen in the 2nd Circuit can be attributed to the fact that 2nd Circuit on average receives a heavier case load than any other circuits. A recent report published by the Administrative Office of the Courts shows that approximately 13.6% of private civil rights claims made in the year of 2013 came from the 2nd Circuit, making the 2nd Circuit the third among all 12 regional Circuits to have the most case load in 2013.³²

³² To estimate the case load of the circuits I used the report provided by the Administrative Office of the Courts. The Administrative Office of the Courts often publishes reports on the case loads of various district and circuit courts and breaks down those case loads into different types of cases. I referred to the report titled “US District Courts—Civil Cases Commenced, by Nature of Suite and District, During the 12-Month Period Ending September 30, 2013” and calculated the number of private civil rights claims that were made for each district in 2013. I used the 2013 measure for convenience and also because there is little reason to believe that the case load for the districts would vary significantly from year to year. I chose to look at private civil rights claims specifically because it serves as the best proxy I could find for employment discrimination cases.

Compared to the other circuits in my population, the DC Circuit also has a large number of cases. It contributes to approximately 12% of the cases in my population making it the second most common circuit to deal with cases having to do with the admissibility of expert evidence in employment discrimination cases. Interestingly, a similar explanation for the large number of cases found in the 2nd Circuit cannot be provided for the DC Circuit, as the Administrative Office of Courts shows that less than 1% of private civil rights claims made in 2013 came from the DC Circuit. This makes the DC Circuit the court that has the least case load in terms of private civil rights claims in 2013.

Plaintiffs

With the exception of three male plaintiffs who filed for Sex discrimination, the majority of plaintiffs who filed for Sex discrimination are females. Similarly, with the exception of three Hispanic plaintiffs who filed for Race discrimination, the majority of plaintiffs who filed for Race discrimination are African Americans. The racial composition of the plaintiffs in my total population includes African American and Hispanic plaintiffs as well as two cases of non-white plaintiffs whose race were not specified.³³

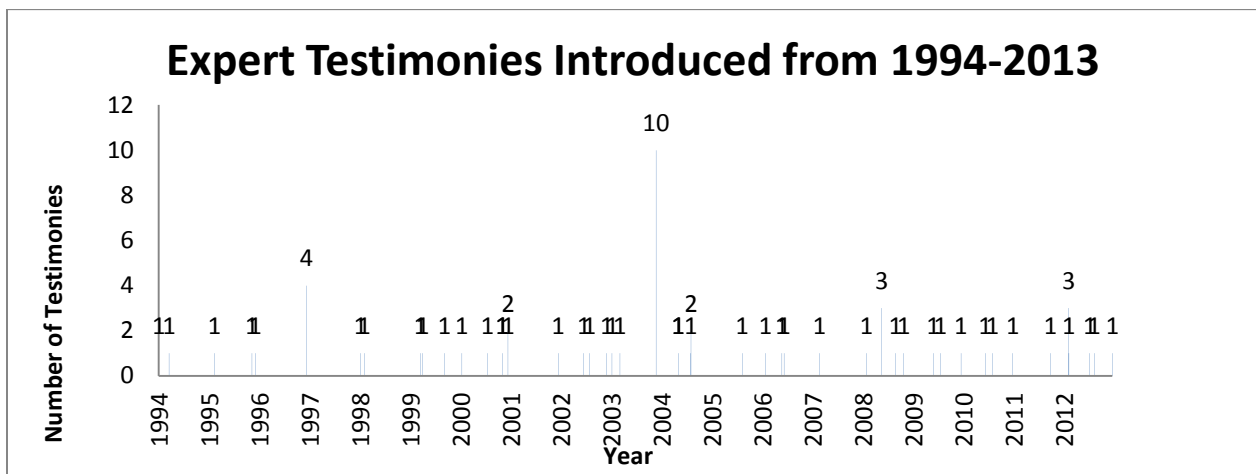
There are no cases where the plaintiff is a public interest organization, but there are seven cases where the plaintiff is a government agency. The U.S. Equal Employment Opportunity Commission served as the plaintiff in five of these cases. The majority of the plaintiffs were neither government agencies nor public interest organizations.

Expert Testimonies across Time

³³ As opinions often did not state the gender of plaintiffs in cases where plaintiffs filed for Race discrimination, and similarly, opinions often did not reveal the race of plaintiffs in cases where plaintiffs filed for gender discrimination, I did not code for the sex of plaintiffs who filed for Race discrimination and the race of plaintiffs who filed for Sex discrimination. As a result, my data captures only the race and gender of the plaintiff's population who filed for either Race or Sex discrimination claims. I also did not code for the gender and race of plaintiffs who did not file for Race or Sex discrimination claims.

There doesn't seem to be any notable trends in the number of expert testimonies that are being introduced to the court over time. The average number of experts introduced by the plaintiff's side is approximately 1 expert per case. It is clear however from the following graph that there are several cases that involve more than one expert.

The chart below displays the number of expert testimonies that were introduced from the plaintiff's side in employment discrimination cases dealing with the admission of expert testimonies from 1994 to 2013.

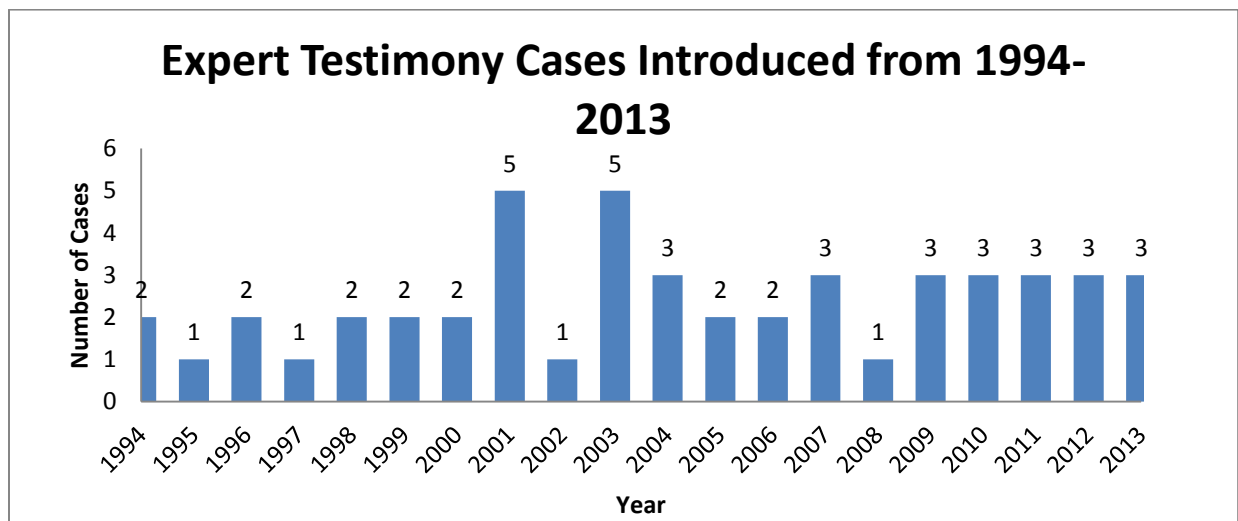


There appears to be no notable trends in the graph, but there are several peaks in the graph where the number of experts introduced far exceeds the average. These cases occur in 1997, 2004, 2009, and 2011. Four expert testimonies were introduced in *Butler v. Home Depot, Inc.* in 1997, 10 expert testimonies were introduced in 2004 in the case of *E.E.O.C. v. Morgan Stanley & Co.*, three expert testimonies were introduced in 2009 in the case of *Tuli v. Brigham & Women's Hosp., Inc.*, and three expert testimonies were introduced in 2012 in the case of *Ellis v. Costco Wholesale Corp.* Of these, the clear outlier is *E.E.O.C. v. Morgan Stanley & Co (2004)* which involved a large scale class action case where multiple experts testified on both sides.

A look at the number of employment discrimination cases dealing with the admissibility of expert testimony to see if there are any trends in the introduction of such cases to the court

across time, showed that the average number of cases introduced concerning the admission of expert testimonies in employment discrimination cases per year is 2.45.

The graph below seems to indicate that such cases became more frequent from 2003 to 2013 as it increased from approximately two of such cases per year to three of such cases per year. On the whole, it appears that the average number of cases dealing with the admissibility of expert testimonies in employment discrimination cases increased over time.



Expert Testimony Admission Rates

The majority of the expert testimonies in my population were admitted with an approximately 63% admission rate.³⁴ Of these cases, 27% of the expert testimonies were struck down. This is consistent with legal scholarship which supports the fact that the *Daubert* decision in application has led to the rise in the exclusion of expert testimony in court.³⁵ At least 10% of cases regarding expert testimony were partially admitted. The Advisory Committee Notes

³⁴ This graph is based on the admissibility of all experts in a case per case. Many cases involved determining the admissibility of several experts. Each expert was weighed separately in each case. The admission was based on the admissibility of each expert's testimony in a case.

³⁵ According to a 2002 RAND study, post *Daubert*, the percentage of expert testimony by scientists that was excluded from the courtroom rose significantly.

accompanying Rule 702 observe that “the rejection of expert testimony is the exception rather than the rule”, this seems to corroborate with my observations here.

Race and Gender Claims and the Admissibility of Expert Testimony

As race and gender discrimination claims are the most prevalent claims in employment discrimination cases, it is interesting to look at whether race and gender claims impact the admissibility of expert testimony. The Chi-Square Test of Independence was used to determine if there is a relationship between the type of claims plaintiffs’ make and the admissibility of expert testimony.³⁶ The null hypothesis is that there is no relationship between the type of discrimination claim that a plaintiff makes and the admissibility of expert evidence. My alternate hypothesis is that there is a relationship between the type discrimination claim a plaintiff makes and the admissibility of expert testimony. I found $X^2(1, N=67) = 5.435788, p < .05$. Because 5.435788 is greater than 3.841 (for $\alpha=0.05$), the H_0 that there is no relationship between the type of discrimination claim that is made and the admissibility of expert evidence is rejected.

	Race claims	Gender claims
Plaintiff’s evidence admitted	29	13
Plaintiff’s evidence not admitted	10	15

³⁶ To conduct the Chi-Square Test of Independence I had to meet a series of assumptions. The chi-squared test, when used with the standard approximation that a chi-squared distribution is applicable, has to follow the following assumptions: (1) simple random sample (2) sample size (3) expected cell count (4) independence. While the cases in my population does not meet the simple random sample assumption, as the data is not randomly sampled from a fixed distribution, it can be generalized to a bigger population of cases involving the admissibility of expert testimony. In addition, as seen in my table, several cells fall short of the expected cell count. Traditional practice requires 5 or more to be the minimum expected cell count for each cell in a table. To achieve this in my case, I combined Plaintiff’s partially admitted evidence column with Plaintiff’s evidence excluded column into the column called Plaintiff’s evidence not admitted.

A simple calculation shows that the admission rate for cases that concern a race claim is approximately 74.358 % as opposed to an admission rate of 46.4286 % for cases that concern gender claims. My finding seems to indicate that the cases that make race claims were more likely to have expert evidence admitted than those that made gender claims.

Admission with Regard to the Existence of an Expert from the Defendant's Side

Another factor I chose to look at is the admissibility of expert testimony with regard to whether the defendant hired his own expert in the case. The assumption is that cases where the defendant brings his or her own expert are less likely to admit the plaintiff's evidence. This reasoning comes from the idea that defendants who have access to their own experts are more likely to have access to other resources that will help them win the case. This is especially true in employment discrimination cases where defendants are far more resourceful and better funded than the plaintiffs.³⁷

To see if there exists a relationship between the existence of an expert on the defendant's side and the admissibility of expert testimony, I conduct the Chi-Square Test of Independence.³⁸ The null hypothesis is that there is no relationship between the introduction of an expert from the defendant's side and the admissibility of expert testimony. The alternate hypothesis is that there is a relationship between the existence of expert testimony by the defendant and the admissibility of expert testimony. The $X^2(1, N=67) = 2.9, p < .05$. As the Chi-Square statistic, 2.9, is smaller than 3.841 (for $\alpha=0.05$), the H_0 that there is no relationship between the introduction of an expert by the defendant and the admissibility of expert testimony fails to be rejected.

³⁷ Galanter, Marc (1974) "Why the 'Haves' Come Out Ahead: Speculations on the Limits of Legal Change," 9 Law & Society Rev. 95-160.

³⁸ Here, similar to the previous case where the Chi-Square test was used, not all the assumptions for the Chi-Square Independence Test were met. To achieve minimum expected cell count, I combined Plaintiff's partially admitted evidence column with Plaintiff's evidence excluded column into the column called Plaintiff's evidence not admitted.

	Defendant has own expert	Defendant does not have own expert
Plaintiff's evidence admitted	29	13
Plaintiff's evidence not admitted	12	13

For defendants who had their own experts, I found that such cases had an expert testimony admission rate of approximately 70 %. On the other hand 50 % of the cases were admitted when the defendants did not have their own expert. This result seems contrary to the intuition that expert testimony admission rates should be lower in cases where the defendant has his own expert than when the defendant has no expert at all.

One explanation for this phenomenon could be that the introduction of experts from both sides leads to a more contentious debate. In this sort of situation, the judge may find an even harder time of determining what is the best course of action in terms of admissibility of evidence. This might make the judge more inclined to admit the evidence and let the experts battle it out in front of the jury.

Reasons for Exclusion of Experts

As established in *Daubert*, judges determine the admissibility of expert testimonies based on a two-pronged test of relevance and reliability. According to the Daubert standard, relevance is determined by whether the testimony relates to the matters that the court needs to address at hand. The relevance standard is not met if the jury is capable of comprehending the primary facts and of drawing correct conclusions from them as are expert witnesses who possess special or

peculiar training. In addition, expert testimony is relevant if it can help jury make useful decisions about the case. If it does not help the jury to this end, the evidence should be excluded.

Daubert also gave a clear outline of what constitutes as reliable testimony. To be reliable, the testimony must meet the following non-exhaustive list of criteria, including: (1) whether the theory or technique on which the expert relies has been or could be tested; (2) whether the theory or technique has been subjected to peer review and publication; (3) the known or potential rate of error of the technique or theory when applied; (4) the existence and maintenance of standards controlling the technique's operation; and (5) whether the theory or technique has been generally accepted in the scientific community. This list is however is not exhaustive as the test of reliability is meant to be a “flexible” one depending on the “nature of the issue, the expert's particular expertise, and the subject of his testimony”.³⁹

I looked at the judicial reasoning of judges in their determination to exclude or partially exclude expert testimony. I looked at cases where expert testimonies were struck and took note of the criteria judges found the testimonies to have failed to meet.⁴⁰

There are a total of 20 cases that dealt with expert testimony that were either excluded or partially excluded. On average it seems that there is an equal number of cases where judges cited relevance or reliability as the criteria for which the expert testimony failed to meet.

Reason Struck	Count
Relevance	6

³⁹ *Daubert*, 509 U.S. at 593-94, 113 S.Ct. 2786;

⁴⁰ I determined if the exclusion of evidence was based solely on relevance or reliability and what types of relevance or reliability concerns the judicial reasons were. I counted more than one reason for each case if there were several reasons that were given by the court.

Reliability	7
Relevance and Reliability	7

I also examined the specific reasons that judges cited under the relevance and reliability concerns they used to justify exclusion of expert testimony. For relevance, the most common reason for citing this criterion was concern that the testimony would not assist the jury. There were ten cases in my sample that cited this reason for exclusion of evidence.

Judges' determination of evidence that does not assist the jury varies from the fact that judges believe that the evidence presented is "within the competency of the average juror", that "the jury can make without assistance from an expert", or that "it relates to matters of common sense". Another less common reason is fear that the testimony offered by the expert would mislead and confuse the jury because that information might be prejudicial and that the testimony has no bearing on any issue in the case. This is based on the standard definition of relevance and occurs when judges find that the evidence introduced does not pertain to the legal question at hand.

Relevance	Count
Does not assist the jury	10
Has no bearing on any issue in the case	2

The determination of relevance is a much more subjective one as compared to determination of reliability. Here, the judge decides what constitutes of evidence such that would be helpful to the jury. This can effectively vary from judge to judge and be influenced by his or

her preconceived notions about the case. Legal scholarship speculates that judges may have preconceived distrust of social science evidence of the type introduced in *Wal-Mart Stores, Inc. vs. Dukes* (2011). It is easy for judges to exclude social science evidence based on the reason that the evidence is not relevant to the discussion at hand. In *Collier v. Bradley University* (2000) for example, Judge Mihm states the following reason for striking down the expert's testimony:

the social psychologist would not be permitted to give expert testimony, it is fundamentally unhelpful to the jury since Collier will be able to tell the jury when she testifies how Defendants alleged conduct made her feel. She does not need the imprimatur of an expert witness to express to the jury how she felt when she worked for Bradley or how she perceived Defendants' conduct.

This statement by Judge Mihm undermines the role that social psychological evidence plays in the courtroom and supports the view that judges hold a skeptical attitude towards social science evidence. Subjective attitudes towards the use of social science evidence can easily creep into the courtroom to exclude valuable information that would otherwise be helpful to the jury. The Plaintiff might want to include such information to bolster her credibility and validate her perceptions. The expert's testimony can effectively add weight to the plaintiff's testimony and help convince the jury the validity of her testimony.

Reliability

As the court has established a much more diverse set of criteria to determine reliability, I also found a more diverse range of judicial reasons that judges gave for reliability concerns. The most common reason is methodological flaw which includes errors produced by faulty assumptions, misuse of datasets, and problems with the analysis of these cases. An example of a methodological flaw can be seen in *Ellis v. Costco Wholesale Corp*(2011) where the judge writes

that the expert had improperly “truncated the time period analyzed in order to reach the conclusion that a statistically significant disparity exists in the average number of years men and women take to be promoted from Senior Staff positions to AGM positions”.

Experts’ qualifications were another major reason that is brought up by the judges’ reasoning. This includes concern with the fact that the expert is not qualified or is incompetent to address the issue at hand, that he or she has no particular expertise that would qualify, or that not enough information was provided about her qualifications. In *Bazile v. City of New York* (2012), the court found that the expert “has no particular expertise that would qualify to assess whether a discriminatory animus motivated the NYPD in this case. Therefore, his testimony is not only based on subjective belief, but also does not fit the facts of the case.”

Then there are concerns that the evidence has not been sufficiently proven or that validation of the evidence is lacking or concerns with the control of the operations of the opinions. These reasons however are much less common.

It is understandable that if judges do not understand scientific evidence very well they would look more towards the qualification criteria in determining the admissibility of expert testimony. However it does not seem that judges are uncomfortable with detecting methodological flaws in the cases as methodological flaws are one of the most commonly cited reasons for exclusion of expert testimony.

Despite of their apparent comfort in identifying methodological flaws in experts’ testimonies, it is unclear whether judges have clear faculties over the operation of such evidence. It is difficult to tell from the opinions alone whether judges are making the correct decision in their understanding of scientific methodologies that are being presented to the court. Nonetheless judges apparent comfort in identifying methodological flaws in cases deviates from the study by

Kovera & McAuliff in 2000 which found that methodological quality influenced judges' ratings of scientific quality and admissibility decisions very little to not at all.⁴¹

Reliability	Count
Qualifications	5
Methodological flaw	7
Not sufficiently proven	2
Not supported by appropriate validation	3
No standards which control the operation of his opinions	1

Statistical Evidence and Legal Theory

Typically, social science evidence makes its way in discrimination cases through disparate impact, systematic disparate treatment, and individual disparate treatment theories of discrimination under Title VII of the United States Civil Rights Act of 1964.

Disparate Impact

Disparate impact theory holds that employment practices may be considered discriminatory and illegal if they have a disproportionate "adverse impact" on persons along the lines of a protected trait. Statistics and social science is often used to prove a prima facie case of disparate impact.

Systemic Disparate Treatment Claims

In systemic disparate treatment claims or pattern and practice, the plaintiff argues that the proportion of members of the protected class in the employer's workforce is significantly smaller

⁴¹ Kovera MB, McAuliff BD *The effects of peer review and evidence quality on judge evaluations of psychological science: are judges effective gatekeepers?* Journal of Applied Psychology, 85(4), Aug 2000, 574-86.

than the proportion in the relevant labor pool, suggesting discrimination. A "pattern or practice" means that the defendant has a policy of discrimination, even if the policy is not always followed. In this case process might be discriminatory such that there is reason to believe that it is motivated by discrimination.

In class actions or other cases alleging a widespread practice of intentional discrimination, plaintiffs may establish a prima facie case using statistical evidence instead of comparative evidence pertaining to each class member. Plaintiffs can combine the statistical evidence with anecdotal or other evidence to support his theory of discriminatory treatment.

Individual Disparate Treatment Claims

In individual disparate treatment claims, the plaintiff, usually an individual, argues that he was treated differently than similarly situated other workers and he introduces statistical evidence showing disparities in the proportion of protected class members relative to the population.

Analysis

I sorted out these discrimination claims and evaluated the admissibility of expert testimonies along these criteria. While most cases in my sample fell into one of the three categories, there were several cases where none of the three types of evidence applied.

As I looked at the admissibility of expert testimony in relation to these three theories of discrimination, I conduct the Chi-Square Test of Independence⁴² to see if there is a relationship between the type of discrimination claim made and the admissibility of expert evidence. I looked at expert testimony admissibility for each of the three employment discrimination theories. The

⁴² To address the assumption that the expected cell count to be 5 for the Chi-square test, I combined the categories of partial admission and total exclusion into the category of non-admission for the purpose of analysis.

following is a table depicting the distribution of testimonies admitted to the court across different legal discrimination theories that were made.

Type of Evidence	Admitted	Not Admitted
Disparate Impact	17	7
Systemic Disparate Treatment	13	8
Individual Disparate Treatment	10	8

My null hypothesis is that there is no relationship between the type of legal theory plaintiffs make and the admissibility of expert testimonies. My alternate hypothesis is that there is a relationship between the type of legal theory plaintiffs make and the admissibility of expert testimony. I found $X^2(2, N=63) = 1.069972826$, $p < .05$. We therefore fail to reject H_0 that there is no relationship between the introduction of an expert by the defendant and the admissibility of expert testimony because 1.069972826 is smaller than 5.991 (for $\alpha = 0.05$).

Looking at each of the discrimination theories individually I find that the admission rate for expert testimony for disparate impact to be approximately 71%, for systemic disparate treatment to be approximately 62% and individual disparate treatment to be approximately 56%. The high admission rate seen for the admission of disparate impact claims could be due to the fact that the court may find it easier to understand statistical arguments in the context of disparate impact cases. Furthermore, the lower admission rates for systemic and individual disparate treatment claims could be due to the fact that disparate treatment involves questions of intent, and courts are reluctant to infer intent based on general statistics. In contrast, disparate impact theories involve inequality of outcome regardless of intent, and so courts may be more receptive

to statistics in these cases. Judges may simply feel more willing to let in the evidence when the question is whether there was a disparate impact on a protected group.

Wal-Mart and its Legacy

The *Wal-mart v. Dukes* decision in 2011 made a huge impact on the legal community.⁴³ It effectively established a higher standard for expert testimony and what constitutes as expert testimony evidence. In *Wal-Mart* the court rejected the social framework testimony of plaintiff's sociological expert because he could not 'determine with any specificity how regularly stereotypes play a meaningful role in employment decisions at Wal-Mart'. The Court emphasized on the plaintiff's failure to calculate whether 0.5% or 95% of the employment decisions at Wal-Mart might be determined by stereotyped thinking. Being able to acquire this statistic requires individualized evidence of actual discriminatory employment decisions to establish a common practice for purposes of class certification, setting a high bar for class certification as well as social science evidence in general.

I looked at the admission of expert testimony before and after the *Wal-Mart* decision to see if they are quantitatively different in any way. As my population of cases only extends to 2013, I only found seven cases that were decisions made post *Wal-Mart*. I also found that the expert admission rate pre *Wal-Mart* to be approximately .67 percent and the admission rate for expert testimony post Wal-Mart to be slightly lower, at approximately .57.14 %. The tables below display the admissibility of expert testimonies in cases before and after the Wal-Mart decision.

Admissibility of Expert Testimony Pre Wal-Mart Decision

⁴³ *Wal-Mart Vs. Dukes* (2011)

Admitted	Denied in Part and Granted in Part	Denied
28	3	15

Admissibility of Expert Testimony Pre Wal-Mart Decision

Admitted	Denied in Part and Granted in Part	Denied
4	2	2

It is difficult to determine based on this small difference the effect of *Wal-Mart* on the admissibility of expert testimonies in employment discrimination cases due to the limited cases post Walmart that are in my sample. Also interesting to note is that none of these cases post *Wal-Mart* dealt with social framework or sociologists specifically.

V. Limitations

There are many limitations to my study. One major limitation is that the population size is a total of 49 cases. This is extremely small. This limitation made it difficult to apply statistical tests for analysis. Nonetheless, the benefit is that I have the access to the entire population of cases I'm interested in.

There is however a question of how trends found here can be generalized to apply in other settings such as other employment discrimination cases or even the court's treatment of expert testimony evidence in other legal areas. There is reason to believe that the effects found in my study might be different for cases from other legal areas. A different pattern has emerged in criminal cases. In criminal cases, the prosecution has the burden of proof and uses a host of forensic science methods as evidence to prove their case, but, *Daubert* motions are rarely made by criminal defendants and when they do, they lose a majority of the challenges.⁴⁴

⁴⁴ Risinger, D. Michael (2000). "Navigating Expert Reliability: Are Criminal Standards of Certainty Being Left on the Dock?". *Albany Law Review* 64: 99. [ISSN 0002-4678](https://www.albany.edu/lawreview/)

Another limitation is time. The cases span only a period of approximately 21 years. This is a limited time frame and it is difficult to see any trends in behavior with simply 49 cases in 21 years. It might be interesting in a future study to do a comparison of pre and post *Daubert* decision making and seeing if there are any significant differences in the way judges' deal with expert testimonies. One other significant limitation is the difficulty of coding cases. The coding scheme is subject and bias can creep in especially with regard to coding statistical significance and trying to code legal reasoning behind the cases.

VI. Conclusion and Further Questions

Through my analysis I was able to uncover evidence supporting my hypothesis that judges are reluctant to admit social science evidence in the courtroom. On the other hand, my analysis also suggests that judges are becoming more comfortable with the use of social science evidence and methodologies and are becoming more discerning their application of the Daubert Standard. I believe that despite of the challenges in interpreting social science evidence in the context of law, it is possible for law and social science evidence to work side by side. I believe that despite of different emphasis on precedent versus change, in the long run, law and science will be able to agree on certain standards of quality control. Such standards are something which the Court are working on developing and as a part of this process, judges must embrace the use of social science evidence in court and work on developing their understanding of different social science and statistical methodologies and models.

In the future, some further research can be done to further examine judge's understanding of social science evidence and statistical methods. An experiment or survey might be a better tool for capturing judge's understanding and attitudes towards social science evidence than judicial opinions which only serve as a selective snapshot of what took place in the courtroom.

VII. Bibliography

- Bartlett, K.T. (2009). *Making Good on Good Intentions: The Critical Role of Motivation In Reducing Implicit Workplace Discrimination*. Virginia Law Review, 95, 1893-1972.
- Bernstein, David E. *Rehabilitating Lochner: Defending Individual Rights against Progressive Reform*. Chapter 4. Chicago: University of Chicago Press. (2011): 195.
- Blumenthal, Jeremy, Cheng Edward, Mnookin Jennifer, *Murphy Erin, and Sanders Joseph*. *“Modern Scientific Evidence: The Law and Science of Expert Testimony*. West/Thomson Publishing Co. 2012-2013. Print.
- Chesler, Mark, Joseph Sanders, and Debra Kalmuss. *Social Science in Court*. Madison: The University of Wisconsin Press, 1988. Print.
- Dixon, L, Gill B. *Changes in the Standards for Admitting Expert Evidence in Federal Civil Cases Since the Daubert Decision*. RAND Institute for Civil Justice. 2002.
- Duizend, Richard, and Saks Michael. *The Use of Scientific Evidence in Litigation*. Williamsburg: National Center for State Courts, 1983. Print.
- Edward K. Cheng, David L. Faigman, Michael J. Saks, and Joseph Sanders. *Modern Scientific Evidence: The Law and Science of Expert Testimony*. Eagan, MN: Thomson/West, 2006.
- Expert Report of William T. Bielby, *Betty Dukes, et al. v. Wal-Mart Stores, Inc.* Feb. 3, 2003
- Faigman, David. Evidentiary Incommensurability: A Preliminary Exploration of the Problem of Reasoning from General Scientific Data to Individualized Legal Decision-Making*, 75 BROOK. L. REV. 1115, 2010.
- Eugene Borgida, and Susan T. Fiske “Beyond Common Sense: Psychological Science in the Courtroom” Blackwell Publishing Ltd. 2008.

- Faigman, David L. (2000) "LECTURE: Legal Alchemy: The Use and Misuse of Science in the Law," *Yale Journal of Law and Technology*: Vol. 2: Iss. 1, Article 3.
- Faigman, David, Nilanjana Dasgupta, and Ridgeway Cecilia. *A Matter of Fit: The Law of Discrimination and the Science of Implicit Bias*, 59 HASTINGS L.J. 1389, 2008.
- Faigman, David. *Evidentiary Incommensurability: A Preliminary Exploration of the Problem of Reasoning from General Scientific Data to Individualized Legal Decision-Making*, 75 BROOK. L. REV. 1115, 2010.
- Faigman, David. *Laboratory of Justice: The Supreme Court's 200-Year Struggle to Integrate Science and the Law*. Henry Holt & Co. 2011. Print.
- Faigman, David. *Alchemy: The Use and Misuse of Science in the Law*. W.H. Freeman & Co. 1999. Print
- Fiske, S. T., Bersoff, D. N., Borgida, E., Deaux, K., & Heilman, M. E. (1991). *Social Science Research on Trial*. *American Psychologist*, 46, 1049–1060.
- Galanter, Marc (1974) "Why the 'Haves' Come Out Ahead: Speculations on the Limits of Legal Change," 9 *Law & Society Rev.* 95-160.
- Gatowski, Sophia I.; Dobbin, Shirley A.; Richardson, James T.; Ginsburg, Gerald P.; Merlino, Mara L.; Dahir, Veronica *Asking the gatekeepers: A national survey of judges on judging expert evidence in a post-Daubert world*. *Law and Human Behavior*, Vol 25(5), Oct 2001, 433-458
- Hart, M , & Secunda, P. (2009). *A Matter of Context: Social Framework Evidence in Employment Discrimination Class Actions*. *Fordham Law Review*, 78, 37-70.
- In E. Borgida & S.T. Fiske (eds.), *Beyond Common sense: Psychological Science in the Courtroom*. (pp. 127-156). Oxford: Blackwell Publishing.

- Kadane, Joseph. *Statistics in the Law*. Oxford: Oxford University Press, 2008. Print.
- Kovera MB, McAuliff BD *The effects of peer review and evidence quality on judge evaluations of psychological science: are judges effective gatekeepers?* *Journal of Applied Psychology*, 85(4), Aug 2000, 574-86.
- Krieger & Fiske *Behavioral realism in employment discrimination law: Implicit bias and Disparate Treatment*. *California Law Review*, 94(4), 997-1062
- Melnick, R. (2005). "A *Daubert* Motion: A Legal Strategy to Exclude Essential Scientific Evidence in Toxic Tort Litigation". *American Journal of Public Health* 95 (S1): S30–S34.
- Michelson, Stephan. *The Expert: The Statistical Analyst in Litigation*. Longbranch Research Associates, 2006. Print.
- Monahan, J., Walker, L., & Mitchell, G. (2009). *The Limits of Social Framework Evidence. Law, Probability, and Risk*, 8(4), 323-351.
- Monahan, John, and Walker Laurens. *Social Science in Law: Cases and Materials*. Foundation Pr; 4th edition. 1998.
- Risinger, D. Michael (2000). "Navigating Expert Reliability: Are Criminal Standards of Certainty Being Left on the Dock?". *Albany Law Review* 64: 99. ISSN 0002-4678
- Roth, Andrea. Safety in Numbers?: Deciding When DNA Alone Is Enough To Convict, 85 *N.Y.U. L. Rev.* 1130 (2010)
- Homes, Oliver. "The Path of the Law" 10 *Harvard Law Review* 457 (1897)
- Woodward, James, "Scientific Explanation", *The Stanford Encyclopedia of Philosophy* (Winter 2011 Edition), Edward N. Zalta (ed.), URL = [<http://plato.stanford.edu/archives/win2011/entries/scientific-explanation/>](http://plato.stanford.edu/archives/win2011/entries/scientific-explanation/).

Saks, Michael, Kaye, David, and Sanders Joseph. *Annotated Reference Manual on Scientific Evidence*. West Publishing Co. 2004.

Tversky, Amos, and Kahneman, Daniel. "Belief in the Law of Small Numbers". Hebrew University of Jerusalem Psychological Bulletin, 1971, Vol. 76, No. 2. 105-110.