

CONSTITUTIONAL FACT FINDING AND THE APPROPRIATE USE
OF EMPIRICAL DATA IN CONSTITUTIONAL LAW

by
*Shawn Kolitch**

In evaluating the constitutionality of a law, the Supreme Court often considers the purpose and effects of the law. This is the case, for instance, in the areas of interstate commerce, equal protection, and substantive fundamental rights, among others. In such cases, empirical data, such as data compiled by social scientists, may be highly relevant to the Court's analysis. However, the Court has been inconsistent in its consideration of empirical data, often misinterpreting available data, and frequently making assertions of fact without any empirical support at all. In some cases, these unsupported assertions appear contrary to both common sense and existing empirical studies, yet they can have a profound impact on the law.

This paper considers the role of empirical data in constitutional law, and proposes that both lack of judicial candor and fundamental tensions between legal and scientific principles contribute to the Supreme Court's inconsistent and often flawed methods of considering empirical data. After examining in detail the role of empirical data in the Court's criminal procedure jurisprudence, the author concludes that 1) the Court should clearly articulate the areas of law in which it believes empirical data is relevant; 2) the Court should consider only peer-reviewed empirical studies and should have access to a scientific advisory committee to help evaluate the reliability of studies under consideration; and 3) a selective relaxation of the Court's stare decisis doctrine is an appropriate mechanism for overcoming the inherent conflict between the scientific method—which requires an evolving understanding of empirical data—and the judicial preference for finality.

I.	INTRODUCTION.....	674
II.	THE RELEVANCE OF EMPIRICAL DATA TO CONSTITUTIONAL INTERPRETATION.....	678
	A. <i>Constitutional Structure and Inherent Irrelevance</i>	678
	1. <i>Separation of Powers</i>	678
	2. <i>Federalism</i>	679

* J.D. 2006, Lewis & Clark Law School; Ph.D. 1996, University of California; M.S. 1990, B.S. 1989 *with honors and distinction*, Columbia University. The author wishes to thank Professor Stephen Kanter for his valuable feedback regarding the development of this paper.

B.	<i>Levels of Scrutiny and Effective Irrelevance</i>	680
1.	<i>Facial Economic Discrimination</i>	681
2.	<i>Facial Discrimination of a Protected Class and Virtual Per Se Unconstitutionality</i>	682
3.	<i>Low-Level Equal Protection Scrutiny</i>	683
C.	<i>Non-Deterministic Scrutiny and Relevant Evidence of Purpose and Effects</i>	683
1.	<i>Facially Neutral Economic Regulation</i>	683
2.	<i>Heightened Equal Protection Scrutiny</i>	684
3.	<i>Substantive Fundamental Rights</i>	685
III.	EMPIRICAL EVIDENCE IN THE LAW OF CRIMINAL PROCEDURE	686
A.	<i>The Exclusionary Rule and Unsupported Assertions</i>	686
1.	<i>The Dual Purposes of the Exclusionary Rule</i>	686
2.	<i>Empirical Evidence of Deterrence as a Straw Man</i>	687
B.	<i>Sixth Amendment Criminal Jury Size Cases and the Tensions Between Legal and Scientific Principles</i>	689
1.	<i>Anecdotal Versus Experimental Evidence</i>	690
2.	<i>Resolution Versus Reliability</i>	692
3.	<i>Revolution Versus Finality</i>	694
4.	<i>Postscript</i>	695
IV.	RESOLVING THE TENSIONS INHERENT IN CONSTITUTIONAL FACT FINDING	696

I. INTRODUCTION

In a legal controversy, a fact-finder must determine relevant facts to some appropriate standard of proof,¹ and then apply the law to those facts. However, unique problems arise when the relevant facts are statistical, rather than particularized in nature. Whereas particularized facts may be proven with highly specific evidence, such as a proverbial “smoking gun” or an eyewitness account, statistical facts generally must be proven with empirical evidence,

¹ Specifically, in criminal cases, the prosecutor must prove that the defendant committed all elements of the accused crime “beyond a reasonable doubt.” In civil cases, the plaintiff typically must prove it was “more probable than not” that the defendant broke the law in the asserted manner, a standard also known as proving the case to a “preponderance of the evidence.” A third and intermediate standard, “substantially more probable than not” or “clear and convincing,” applies in some civil matters, typically those in which a presumption of some sort must be overcome. For example, an accused patent infringer who asserts invalidity of the patent as a defense must prove invalidity by clear and convincing evidence to overcome a presumption that the patent is valid. *Geneva Pharm., Inc. v. Glaxosmithkline PLC*, 349 F.3d 1373, 1377 (Fed. Cir. 2003). Other examples include involuntary civil commitments, and termination of parental rights. *See Santosky v. Kramer*, 455 U.S. 745, 753, 766–67 (1982) (requiring a clear and convincing evidence standard for termination of parental rights to overcome a presumption that the best home for the child is with the parents); *Addington v. Texas*, 441 U.S. 418, 426–27 (1979) (requiring clear and convincing evidence in an involuntary commitment hearing to overcome a presumption that all persons are legally competent).

such as data from social scientific studies. As a result, statistical facts demand special consideration for at least two reasons. First, the relevance and scope of legally admissible empirical evidence must be determined. Although in principle, a legal system could limit this scope to evidence having implications only for a particular case or controversy, the rules of evidence in the United States allow a party to introduce more general empirical evidence at trial,² and the implications of admitting such general evidence may extend beyond the case at issue. Even more broadly, no theoretical or legal obstacle prevents courts from considering empirical evidence when interpreting the U.S. Constitution. In fact, the Supreme Court already occasionally uses such evidence to justify its holdings.³ Second, once the scope of admissible empirical evidence is established, the courts still must select acceptable methods and standards for proving statistical facts. Neither the Constitution nor the Supreme Court's past opinions provide clear guidance in these matters.⁴

In general, statistical facts may be classified in one of three categories. First, *case-specific statistical facts* are those that apply only to a particular case or controversy, and the broader application of these facts usually would be inappropriate, if not nonsensical. For example, in an action for trademark infringement, a plaintiff may seek to prove that due to the similarity between the plaintiff's registered trademark and an allegedly infringing mark, a "likelihood of confusion" exists among consumers about the source of particular goods or services.⁵ In this case, the probabilistic nature of the word "likelihood" suggests that at least some of the proof may come in the form of statistical data. Indeed, an acceptable—and commonplace—method of proof in such cases is to present the results of consumer surveys indicating some degree of actual consumer confusion stemming from the conflicting trademarks. However, although the fact that a likelihood of confusion does or does not exist among consumers with respect to specific salable items is crucial for determining the outcome of a trademark infringement suit, this fact has no

² The Federal Rules of Evidence define admissible relevant evidence as "evidence having any tendency to make the existence of any fact that is of consequence to the determination of the action more probable or less probable than it would be without the evidence." FED. R. EVID. 401. Although this standard applies at trial, and appellate courts generally do not hear new evidence, cases are commonly remanded to a trial court with specific instructions to develop the record further in a certain regard. In this manner, appellate courts effectively can direct the introduction of empirical evidence for appellate review.

³ This may occur through remand, as discussed *supra* note 2, or it may occur through the use of a so-called Brandeis brief, a document advanced at the appellate level to persuade the court and that may include, *inter alia*, economic and social scientific studies, reports of public committees, and evaluations of issues by scientific experts. The term "Brandeis brief" has its origins in a brief submitted to the Supreme Court by Louis D. Brandeis in *Muller v. Oregon*, 208 U.S. 412 (1908). Brandeis himself later became an Associate Justice of the Supreme Court.

⁴ The Constitution makes no references to evidentiary questions or to methods of constitutional interpretation. As discussed *infra* Part III, the Court's holdings to date are unclear if not inconsistent.

⁵ The "likelihood of confusion" is a staple of trademark law that has its federal origins in section 43(a) of the Lanham Trademark Act. 15 U.S.C. § 1125(a) (2000).

particular relevance beyond the case at issue. Thus, a case-specific statistical fact of this nature would be unlikely to have broader applicability even within a specific realm of law (in this case, trademark law), and almost definitely would be irrelevant to questions of either statutory or constitutional interpretation.

A second and more general class of statistical facts includes *background statistical facts* commonly used to support an inference, particularly with regard to the reliability of another piece of evidence. For instance, in a child abuse case, an expert psychologist might testify that a particular behavior pattern exhibited by a child is indicative of past abuse of a certain type. Here, the expert may not be making a factual assertion regarding the particular child involved in the case, and in fact may not even have examined the child. Rather, the expert is asserting the general proposition that certain behavior is statistically likely to be linked to past abuse, based on the expert's experience with a large number of similar cases. Witnesses such as the child, the parents, or another expert may be used to establish the case-specific fact that the child does indeed exhibit that behavior. The background statistical fact of the likely link between past abuse and present behavior may have relevance beyond one particular case, to a class of cases involving similar patterns of behavior, and indeed the same expert may give essentially the same testimony in many such cases. However, like case-specific statistical facts, background statistical facts are designed to support a specific proposition mandating the outcome of the case at issue, rather than a general one mandating a change in the law, and thus are unlikely to be relevant to questions of constitutional interpretation.

A third category of statistical facts are those I will term *constitutional statistical facts*. These are facts based on empirical data and are relevant to constitutional interpretation. In general, constitutional statistical facts support a proposition of the nature "we should interpret the Constitution in manner Y, at least in part because statistical fact X is true." For example, in *Williams v. Florida*,⁶ the Court held that a six-person jury in a criminal trial was sufficient to comply with a defendant's Sixth Amendment right to trial by jury, in part because "[w]hat few experiments have occurred—usually in the civil area—indicate that *there is no discernible difference between the results reached by the two different-sized juries.*"⁷ Whether or not there is a discernible difference

⁶ 399 U.S. 78 (1970).

⁷ *Id.* at 101–102 (emphasis added). Note, however, that the ability to reach fair outcomes consistently was only one feature of the jury considered by the Court. Indeed, the Court stated that "the essential feature of a jury obviously lies in the interposition between the accused and his accuser of the commonsense judgment of a group of laymen, and in the community participation and shared responsibility that results from that group's determination of guilt or innocence. The performance of this role is not a function of the particular number of the body that makes up the jury." *Id.* at 100. The Court used similar logic in upholding statutes allowing nonunanimous verdicts in criminal cases. *See, e.g., Apodaca v. Oregon*, 406 U.S. 404, 411 (1972) ("Requiring unanimity would obviously produce hung juries in some situations where nonunanimous juries will convict or acquit. But in either case, the interest of the defendant in having the judgment of his peers interposed between himself and the officers of the State who prosecute and judge him is equally well served."). Whether the interposition of the jury between the accused and the

between the results reached by six-person juries and twelve-person juries in criminal trials is a question of constitutional statistical fact that transcends the specific controversy of *Williams*, and influences how the Court chooses to interpret and apply the Sixth Amendment.

In *Williams*, the Supreme Court asserted a constitutional statistical fact based on “a few experiments . . . in the civil area,”⁸ suggesting a willingness to base such assertions on available empirical data. However, even within the area of Sixth Amendment jurisprudence—and certainly across different areas of law—the Court’s decisions show little consistency with respect to its treatment of constitutional statistical facts. In *Williams*, for example, the Court based its assertion about the equivalence of six- and twelve-person juries on anecdotal evidence rather than scientific data, arguably as a pretense for supporting a conclusion it had already reached in the case.⁹ In subsequent cases revisiting the issue of minimum jury size, the Court first prematurely relied upon subsequently obtained empirical data to uphold *Williams*,¹⁰ and later used more carefully vetted statistical studies to support a legal conclusion seemingly at odds with the factual conclusions suggested by the studies and acknowledged by the Court.¹¹ In short, the Court routinely interprets empirical data in a remarkably unscientific manner. Furthermore, in other areas of the law, the Court frequently makes assertions of constitutional statistical fact with no empirical foundation whatsoever.¹² Yet these assertions can have a profound impact on the law. In this Comment, I address the questions of when and how the Court should rely on empirical data to support an assertion of constitutional statistical fact.

Part II of this Comment discusses the general question of when empirical data is relevant to constitutional interpretation, and concludes that fact-finding supported by statistical empirical data is inherently relevant and vital to a well-grounded decision in many areas. This is true either at the trial court level through the introduction of admissible evidence, or at the appellate level through the introduction of Brandeis briefs and the like. Part III discusses some of the various inconsistent approaches the Court has taken in its constitutional fact finding, focusing on the Court’s exclusionary rule jurisprudence and on the *Williams* line of jury size cases introduced above, and suggests that the Court’s inconsistency is due in large part to a lack of judicial candor and fundamental tensions between the disparate goals of science and the law. Part IV concludes that to resolve these inherent tensions, the Court can and should adopt a form of

accuser has much meaning if the jury is too small or too divided to render a reliable verdict seems questionable.

⁸ *Williams*, 399 U.S. at 101–02.

⁹ See *infra* Part III.B for a more detailed discussion of *Williams* and the subsequent line of cases related to the constitutionally minimum size of a criminal jury.

¹⁰ See *Colgrove v. Battin*, 413 U.S. 149 (1973) (affirming *Williams* and allowing six-person juries in state civil trials as well).

¹¹ See *Ballew v. Georgia*, 435 U.S. 223 (1978) (again affirming *Williams*, but holding five-person criminal juries unconstitutional).

¹² See *infra* Part III.A for an example of this in the context of the Fourth Amendment.

the scientific method to obtain and evaluate empirical data in certain types of cases.

II. THE RELEVANCE OF EMPIRICAL DATA TO CONSTITUTIONAL INTERPRETATION

The Supreme Court need not engage in constitutional fact finding to decide every case, because in many cases—and even in some general areas of constitutional law—empirical data is either inherently or effectively irrelevant. As Justice Marshall famously stated in *Marbury v. Madison*, “[a]n act of congress repugnant to the constitution cannot become a law.”¹³ The Court often has found both legislative acts and common law decisions to be facially inconsistent with the Constitution, in a manner essentially independent of the factual basis of the underlying case. This may occur, for example, when a ruling is based on purely theoretical grounds, or when the level of scrutiny applied by the Court obviates the need for empirical support. On the other hand, cases exist in which empirical data are so relevant as to be essential to a well-grounded decision. This section explores the general question of when constitutional facts are relevant to the Supreme Court’s jurisprudence.

A. *Constitutional Structure and Inherent Irrelevance*

Some areas of constitutional interpretation generally involve purely theoretical considerations, and empirical support therefore may be irrelevant. This may be true, for example, in cases where the structure of the law—rather than its purposes or effects—is at issue. Two categories of cases in which this often is true are those involving separation of powers and federalism issues.

1. *Separation of Powers*

In cases involving the separation of powers within the federal government, the holding typically relies on the basic structure of power outlined by the text of the Constitution, and not on an empirical foundation extending beyond the case-specific facts. In other words, although each case must stem from an underlying controversy to be constitutionally ripe,¹⁴ this underlying controversy in separation of powers cases either is not representative of a large number of similar controversies upon which the Court’s decision will have an effect, or if it is, the Court’s theoretical rationale outweighs such practical considerations. As a result, the Court need not consider statistical empirical evidence regarding the impact of the case.

¹³ 5 U.S. (1 Cranch) 137, 138 (1803).

¹⁴ The ripeness doctrine holds that federal courts generally will not issue advisory opinions. *See, e.g., United Pub. Workers of Am. v. Mitchell*, 330 U.S. 75, 89 (1947) (holding that in a case involving a challenge to a section of the Hatch Act of 1940, the challenger’s “generality of objection [was] really an attack on the political expediency of the Hatch Act, not the presentation of legal issues. It is beyond the competence of courts to render such a decision.”).

For example, in *Plaut v. Spendthrift Farm*,¹⁵ the Court held that Congress could not instruct federal courts to reopen final judgments without violating separation of powers principles, and applied this holding to strike down a provision of the Securities Exchange Act. The Court did not base its conclusion on empirical facts such as the legislative purpose of the Act or its actual effects on federal cases, nor did the Court weigh the potential effects of its holding on other federal statutes. Rather, the Court simply stated that Congress had exceeded its authority by requiring the federal courts to exercise judicial power “in a manner repugnant to the text, structure, and traditions of Article III.”¹⁶ In cases such as this, where the Court finds a purely theoretical constitutional basis for its holding, empirical evidence is inherently irrelevant to the analysis.

2. Federalism

Similarly, federalism cases generally revolve around the division of government power, although here the division is between the federal government and the states rather than between coordinate branches of the federal government. However, as in separation of powers cases, the questions at issue are often theoretical, relating to the structure of the government rather than to the measurable effects of its laws. For example, in *Printz v. United States*,¹⁷ the Court struck down the Brady Act, which required local and state law enforcement officials to collect and submit to the federal government data on those purchasing handguns. The court held that such a requirement was a violation of state sovereignty, because “the Federal Government may not compel the States to implement, by legislation or executive action, federal regulatory programs.”¹⁸ In federalism cases of this type, theoretical considerations alone may be used to support a holding, and empirical considerations may be inherently irrelevant.

On the other hand, federalism cases exist in which empirical data may play an important role. One such category of cases involves the Court’s Eleventh Amendment interpretation, where in making exceptions to the Amendment’s ban on suits against the states, the Court has often looked for empirical evidence of a pattern of abuse in the past actions of the states that would justify revoking their legal immunity. For example, in *Florida Prepaid Postsecondary Education Expense Board v. College Savings Bank*,¹⁹ the Court restricted individuals from bringing patent infringement suits against the states, due to a lack of empirical evidence of state abuses of the patent system. In its opinion, the Court observed that “[t]he Federal Circuit in its opinion identified only eight patent-infringement suits prosecuted against the States in the 110 years between 1880 and 1990.”²⁰ In federalism cases where evidence about the past

¹⁵ 514 U.S. 211 (1995).

¹⁶ *Id.* at 218.

¹⁷ 521 U.S. 898 (1997).

¹⁸ *Id.* at 924.

¹⁹ 527 U.S. 627 (1999).

²⁰ *Id.* at 640.

behavior of the states is important, empirical data may be relevant to the Court's jurisprudence.

B. Levels of Scrutiny and Effective Irrelevance

Many cases exist in which empirical data that might otherwise be relevant becomes effectively irrelevant. This generally occurs because some other consideration supersedes empirical considerations, which are therefore never reached. For example, the Court often strikes down facially discriminatory economic regulations as unconstitutional restrictions on interstate commerce, without considering the actual effects of the law. Similarly, although equal protection jurisprudence primarily relates to the effects of a law on a class of citizens, and empirical data may be highly relevant in determining those effects, the Court's need to rely on empirical data in a given equal protection case often depends on the level of scrutiny the Court finds appropriate in that case. As Justice O'Connor recently stated, "[i]t is unlikely, in my opinion, that any . . . classifications based on stereotypes can survive heightened scrutiny, but under rational scrutiny, a statute may be defended based on generalized classifications unsupported by empirical evidence."²¹

Thus, as Figure 1 illustrates (see below), at one extreme the Court may apply strict scrutiny and strike down a facially discriminatory law, without relying on empirical data regarding the purpose or effects of the law. At the other extreme, in cases where it applies low-level, rational basis scrutiny, the Court may uphold a statute without requiring empirical evidence, because such evidence is generally not required to sustain the classification of an unprotected class. A broad middle ground is occupied by many cases in which the Court's scrutiny is neither so strict nor so relaxed as to avoid practical questions about the purpose and effects of a law.²² Empirical data may be important in these cases.

²¹ *Miller v. Albright*, 523 U.S. 420, 452 (1998) (O'Connor, J., concurring).

²² Note that the two inquiries are analytically separate, because even a law having a completely legitimate purpose which is well-supported by empirical evidence may not have effects corresponding to that purpose. In fact, all laws generally will be overinclusive, underinclusive, or both to some degree. Conversely, a law having an entirely illegitimate purpose may be struck down regardless of its actual effects.

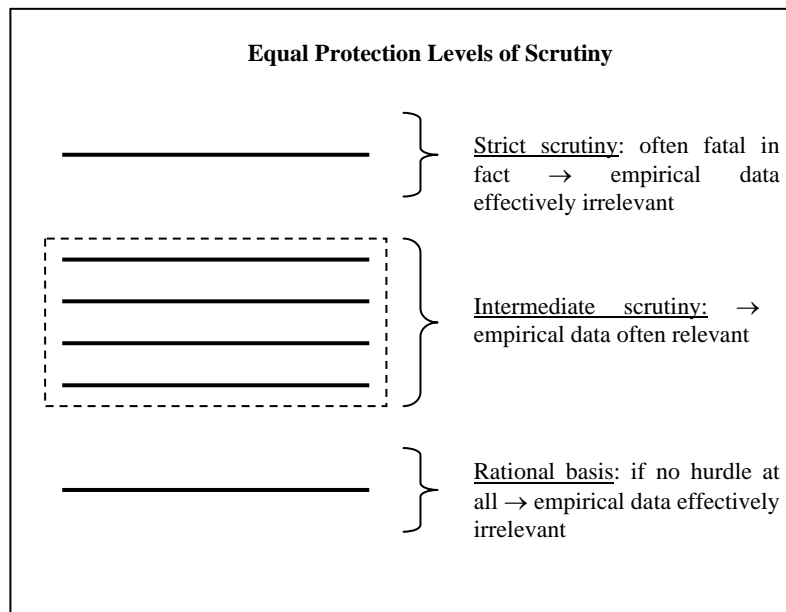


Figure 1

In cases where the level of scrutiny is excessively high or low, the constitutional issues are decided by considerations essentially unrelated to the purpose and effects of a questioned law, and empirical data becomes effectively irrelevant to the Court's jurisprudence. However, in a large number of middle ground cases, in cases where strict scrutiny is not fatal, and in cases of low-level scrutiny with "bite," empirical data plays an important role.

1. Facial Economic Discrimination

In what amounts to an application of strict scrutiny, the Court often has struck down facially discriminatory economic regulations as improper restrictions on interstate commerce, with little or no regard as to the actual purpose or effects of the law, and thus with little or no need to reach empirical questions.²³ An example is the seminal case of *City of Philadelphia v. New Jersey*,²⁴ in which the Court held that a New Jersey statute prohibiting

²³ However, such facially discriminatory regulations are not *per se* unconstitutional. As the Court explained in *Hughes v. Oklahoma*, if a law is facially or effectively discriminatory against interstate commerce, the State has the burden to show both that the law "serves a legitimate local purpose," and that this purpose could not be served as well by available nondiscriminatory means. 441 U.S. 322, 336 (1979). An example of a case where a facially discriminatory state law was upheld under this test is found in *Maine v. Taylor*, 477 U.S. 131 (1986), in which a statute prohibiting importation of baitfish into Maine was upheld because the Court found that the State's purpose of protecting native species could not be served as well by any nondiscriminatory means.

²⁴ 437 U.S. 617 (1978).

importation of garbage collected outside the state was unconstitutional because it was facially discriminatory. In *City of Philadelphia*, the Court stated that “whatever New Jersey’s ultimate purpose, it may not be accomplished by discriminating against articles of commerce coming from outside the State.”²⁵ Similarly, in *C & A Carbone, Inc. v. Town of Clarkstown*,²⁶ the Court held that a Clarkstown ordinance requiring all waste brought into or shipped out of Clarkstown to be sorted in the local transfer station was unconstitutional, as it was facially discriminatory against out-of-state or non-local garbage processors and “would excite those jealousies and retaliatory measures the Constitution was designed to prevent.”²⁷ Thus, an economic regulation which facially distinguishes between intrastate and interstate parties is presumptively invalid and may not survive the Court’s initial scrutiny, regardless of empirical evidence showing that the law has a legitimate legislative purpose and/or positive economic effects.

2. *Facial Discrimination of a Protected Class and Virtual Per Se Unconstitutionality*

Similarly, the Court may strike down a law which is facially discriminatory against a protected class of individuals on equal protection grounds, without considering empirical evidence related to the purpose or effects of the law. Although the Court in these cases is technically applying strict scrutiny, this scrutiny is so frequently fatal to the law that it closely resembles a rule of *per se* unconstitutionality.²⁸ For example, in *Palmore v. Sidoti*,²⁹ the Court held that it was facially discriminatory and unconstitutional for a state court to remove a child from the custody of her mother based on the mother’s cohabitation with a black man, stating that “[a] core purpose of the Fourteenth Amendment was to do away with all governmentally imposed discrimination based on race.”³⁰ Similarly, in *Sugarman v. Dougall*,³¹ the Court applied heightened equal protection scrutiny to a New York statute denying aliens the right to hold positions in the state’s classified competitive civil service, stating that the statute “is neither narrowly confined nor precise in its application”³² and therefore “violates the Fourteenth Amendment’s equal protection guarantee.”³³ In such cases of facial unconstitutionality, the Court typically never reaches questions related to the purpose or effects of the invalid laws, and thus never seeks answers requiring an empirical foundation. Thus, strict scrutiny in the context of equal protection amounts to a rule of virtual *per se* unconstitutionality, rendering empirical support for the Court’s holding

²⁵ *Id.* at 625-27.

²⁶ 511 U.S. 383 (1994).

²⁷ *Id.* at 390.

²⁸ A notable exception is *Korematsu v. U.S.*, 323 U.S. 214 (1944), in which the Court upheld wartime statutes discriminating against Japanese Americans.

²⁹ 466 U.S. 429 (1984).

³⁰ *Id.* at 432.

³¹ 413 U.S. 634 (1973).

³² *Id.* at 643.

³³ *Id.* at 646.

effectively irrelevant even though data related to the purpose and effects of the questioned law might have been relevant if the law had survived initial scrutiny by the Court.

3. *Low-Level Equal Protection Scrutiny*

On the other hand, in equal protection cases where it applies low-level, rational basis scrutiny, the Court often upholds statutes without considering empirical evidence, and has explicitly stated that empirical evidence is not required to sustain the classification of an unprotected class.³⁴ For example, in *FCC v. Beach Communications, Inc.*,³⁵ the Court upheld the Cable Communications Policy Act despite its distinction between facilities that served separately owned and managed buildings, and those that served one or more buildings under common ownership or management for purposes of meeting a franchise requirement. The Court stated that “a legislative choice is not subject to courtroom fact-finding and may be based on rational speculation unsupported by evidence or empirical data.”³⁶ Clearly, if a legislature need not justify its laws with empirical data, the Court has no need to evaluate such data when analyzing the constitutionality of those laws.

C. *Non-Deterministic Scrutiny and Relevant Evidence of Purpose and Effects*

In many cases, initial scrutiny of a law is insufficient either to uphold or invalidate the law on its face. In these cases, as its next step the Court often views the law at least partially, if not primarily, with pragmatic considerations in mind, so that empirical data is inherently relevant. These pragmatic considerations generally relate to the purpose and effects of the questioned law.

1. *Facially Neutral Economic Regulation*

In the context of interstate commerce, a facially neutral law still may be unconstitutional if it has a discriminatory purpose or effect,³⁷ and the Court may consider empirical data in making this determination. In a first step of analysis, the Court may consider data in determining whether the purpose of a statute is legitimate, as in *Southern Pacific Co. v. Arizona ex rel. Sullivan*,³⁸ where the Court struck down an Arizona law barring operation of trains having more than 14 cars on the grounds that there was a lack of evidence supporting the state’s claimed purpose of improved rail safety. In *Southern Pacific Co.*, the Court stated that “[e]xamination of the evidence and the detailed findings makes it clear that . . . such increased danger of accident and personal injury as may result from the greater length of trains is more than offset by the increase in the number of accidents resulting from the larger number of trains when train

³⁴ See *Miller v. Albright*, 523 U.S. 420, 452 (1998).

³⁵ 508 U.S. 307 (1993).

³⁶ *Id.* at 315.

³⁷ See *Minnesota v. Clover Leaf Creamery Co.*, 449 U.S. 456, 471 n.15 (1981) (“A court may find that a state law constitutes ‘economic protectionism’ on proof either of discriminatory effect . . . or of discriminatory purpose.”) (citation omitted).

³⁸ 325 U.S. 761 (1945).

lengths are reduced.”³⁹ In other words, empirical evidence may be relevant to undermine or strengthen the legal legitimacy of a law’s purpose.⁴⁰

More commonly, the Court finds a legitimate purpose, but still may evaluate empirical data to determine whether the actual effects of a law regulating commerce are discriminatory. For example, in *Hunt v. Washington State Apple Advertising Commission*,⁴¹ the Court found that a North Carolina law requiring United States Department of Agriculture (USDA) gradation of apples had the legitimate state purpose of attaining uniform quality standards, but had the discriminatory effect of forcing Washington apple growers to remark their apples for shipment to North Carolina. As one basis for this holding, the Court asserted the empirical fact that “apple brokers and dealers located both inside and outside of North Carolina . . . state their preference, and that of their customers, for apples graded under the Washington, as opposed to the USDA, system[.]”⁴² Thus, empirical facts may be used to evaluate the discriminatory effects of an economic law.

2. Heightened Equal Protection Scrutiny

In many equal protection cases, heightened scrutiny is insufficient to invalidate a law on its face, but demands that the law be supported by a sufficiently compelling purpose and sufficiently nondiscriminatory effects. This is true even in cases invoking strict scrutiny, when the law in question is not invidiously discriminatory, and also in cases involving intermediate levels of scrutiny, such as those involving gender discrimination or discrimination against disabled persons. In such cases, the Court may consider available data as part of its analysis. For example, in upholding the affirmative action policy of the University of Michigan Law School, the Court in *Grutter v. Bollinger*⁴³ applied strict scrutiny, but found that racial diversity among the student body was not an illegitimate purpose because “[i]n addition to the expert studies and reports entered into evidence at trial, numerous studies show that student body diversity promotes learning outcomes, and ‘better prepares students for an increasingly diverse workforce and society, and better prepares them as professionals.’”⁴⁴ Thus, empirical studies may be relevant to equal protection jurisprudence even under a strict scrutiny analysis.

Similarly, in striking down an Oklahoma law prohibiting the sale of 3.2% beer to males under the age of 21 and females under the age of 18 on equal

³⁹ *Id.* at 775.

⁴⁰ On the other hand, some members of the Court have stated in a similar context that there should be no appellate review of such evidence. *See, e.g., Kassel v. Consol. Freightways Corp.*, 450 U.S. 662, 669 (1981) (Brennan, J., concurring) (“[t]he courts are not empowered to second-guess the empirical judgments of lawmakers concerning the utility of legislation.”).

⁴¹ 432 U.S. 333 (1977).

⁴² *Id.* at 351.

⁴³ 539 U.S. 306 (2003).

⁴⁴ *Id.* at 330 (citing Brief for American Educational Research Association et al. as Amici Curiae at 3, *Grutter*, 539 U.S. 306 (No. 02-241)).

protection grounds, the Court in *Craig v. Boren*⁴⁵ analyzed the effects of the law under heightened (mid-level) scrutiny, and stated that “the empirical data submitted by the State accentuate the unfairness of treating all 18–21-year-old males as inferior to their female counterparts.”⁴⁶ Furthermore, despite applying only low-level scrutiny to a zoning ordinance excluding group homes for the mentally retarded from certain zoning districts, the Court in *City of Cleburne v. Cleburne Living Center*⁴⁷ struck down the ordinance, finding it based “on an irrational prejudice against the mentally retarded.”⁴⁸ Thus, in equal protection cases where a law is neither facially acceptable under low-level scrutiny, nor facially invalid under strict scrutiny, the Court has acknowledged the relevance of constitutional statistical facts, and often weighs empirical evidence of such facts in its holdings.

3. *Substantive Fundamental Rights*

In cases implicating fundamental rights, the Court generally applies a form of strict scrutiny and considers, *inter alia*, both the burden the law imposes upon private citizens, and the significance of the state interest. If the burden is direct and substantial, then the state must prove that a truly significant interest is served by the law.⁴⁹ For example, in striking down a dual parental notification provision in a Minnesota law regulating teen abortions, the Court in *Hodgson v. Minnesota*⁵⁰ stated that “[t]he judges who adjudicated over 90% of these petitions testified; none of them identified any positive effects of the law.”⁵¹ In a partial dissent, Justice Marshall argued that the 48-hour waiting requirement upheld by the Court also should have been struck down, stating “I base my conclusion not on my intuition about the needs and attitudes of young women, but on a sizable and impressive collection of empirical data documenting the effects of parental notification statutes and of delaying an abortion.”⁵² These statements from *Hodgson* typify the Court’s responses to questions regarding the effects of a law restricting a fundamental right. In such cases, empirical data often is relevant and may be crucial to a well-supported decision.

Having established that empirical data may be both inherently and effectively relevant to constitutional interpretation, I now turn to a more detailed analysis of some of the various ways in which the Supreme Court has

⁴⁵ 429 U.S. 190 (1976).

⁴⁶ *Id.* at 214.

⁴⁷ 473 U.S. 432 (1985).

⁴⁸ *Id.* at 450.

⁴⁹ *See, e.g., Roe v. Wade*, 410 U.S. 113, 155 (1973) (“Where certain ‘fundamental rights’ are involved, the Court has held that regulation limiting these rights may be justified only by a ‘compelling state interest’”) (citations omitted); *Planned Parenthood of Se. Pa. v. Casey*, 505 U.S. 833, 877 (1992) (“Regulations designed to foster the health of a woman seeking an abortion are valid if they do not constitute an undue burden.”).

⁵⁰ 497 U.S. 417 (1990).

⁵¹ *Id.* at 441.

⁵² *Id.* at 464 (Marshall, J., concurring in part, concurring in the judgment in part, and dissenting in part).

treated empirical evidence, focusing on the Court's use of empirical studies in the area of criminal procedure.

III. EMPIRICAL EVIDENCE IN THE LAW OF CRIMINAL PROCEDURE

Statistical facts are highly relevant in many areas of criminal procedure law, including cases involving Fourth Amendment search and seizure issues (where procedural rules often are designed to deter misconduct by police and other actors), jury selection and other Sixth Amendment issues (which often invoke statistical factual assertions regarding the effects of jury selection practices on trial outcomes), and death penalty and other Eighth Amendment issues (which involve questions about deterrence and public opinion). In these and other areas of criminal procedure, the Supreme Court's jurisprudence provides numerous examples where the Court—either tacitly or explicitly—makes assertions of constitutional statistical fact, generally related to the effects of mandated procedural practices upon law enforcement personnel, criminal defendants, or the lower courts, to support a holding. In this section, I explore the Court's use of factual assertions in criminal procedure cases, and argue that those assertions are often largely unsupported by valid empirical data.

A. *The Exclusionary Rule and Unsupported Assertions*

The Warren Court's expansion of individual constitutional rights, viewed by some as excessively limiting to government action, led to a post-Warren backlash in which those same rights were subsequently curtailed, often on the basis of unsupported assertions of constitutional fact. A case in point is the Court's Fourth Amendment jurisprudence, in which the Warren Court expanded limits on warrantless government searches and seizures in *Katz v. United States*,⁵³ and then eroded those limits in a series of later decisions. In an oft-quoted phrase from his concurrence in *Katz*, Justice Harlan characterized Fourth Amendment limitations as stemming from an individual's right to be free from government intrusions in places where the individual has a "reasonable expectation of privacy."⁵⁴ However, in many of its subsequent holdings limiting the breadth of *Katz*, the Court shifted its perspective to focus on the goal of deterring police misconduct rather than that of protecting individual rights, and frequently relied on unsupported assertions of fact to support its position.

1. *The Dual Purposes of the Exclusionary Rule*

Long before *Katz*, the Supreme Court had created an exclusionary rule requiring evidence obtained in violation of the Fourth Amendment to be excluded from the government's case against the accused.⁵⁵ However, whether

⁵³ 389 U.S. 347 (1967).

⁵⁴ *Id.* at 360 (Harlan, J., concurring).

⁵⁵ See *Weeks v. United States*, 232 U.S. 383 (1914) (adopting the exclusionary rule in federal prosecutions); *Mapp v. Ohio*, 367 U.S. 643 (1961) (extending the exclusionary rule

the rationale for the exclusionary rule is to protect individual rights, to deter government misconduct, or both, has been the subject of much debate. In the seminal case of *Weeks v. United States*,⁵⁶ the Court suggested that individual rights were paramount, stating:

If letters and private documents can thus be seized and held and used in evidence against a citizen accused of an offense, the protection of the Fourth Amendment declaring his right to be secure against such searches and seizures is of no value, and, so far as those thus placed are concerned, might as well be stricken from the Constitution.⁵⁷

However, the Court first suggested in *Elkins v. United States*⁵⁸ that deterrence was the primary purpose of the rule, stating that “[i]ts purpose is to deter—to compel respect for the constitutional guaranty in the only effectively available way—by removing the incentive to disregard it.”⁵⁹ Since deterrence is a measurable effect, empirical data became relevant to the Court’s Fourth Amendment jurisprudence.

2. *Empirical Evidence of Deterrence as a Straw Man*

In *Elkins*, the Court recognized that data regarding the deterrent effect of the exclusionary rule would be difficult to obtain.⁶⁰ However, this difficulty was relatively unimportant at the time, since the goal of the Court in *Elkins* was to restrict rather than to expand government power, by excluding evidence illegally obtained by state actors from federal trials.⁶¹ Only when the Court began to limit the power of the exclusionary rule in the post-Warren era did evidence of deterrence become more important, since the Court was forced to justify narrowing the scope of the Fourth Amendment by focusing on the rule’s lack of a deterrent effect in various circumstances. The fact that no reliable data regarding the deterrent effect of the exclusionary rule existed did not prevent the Court from invoking empirical considerations to support its position. In fact, as the cases discussed below illustrate, the Court used the lack of empirical evidence about deterrence to its advantage.

to state prosecutions); *Wong Sun v. United States*, 371 U.S. 471 (1963) (applying the rule to the fruits of excluded evidence).

⁵⁶ 232 U.S. 383 (1914).

⁵⁷ *Id.* at 393.

⁵⁸ 364 U.S. 206 (1960).

⁵⁹ *Id.* at 217.

⁶⁰ *See id.* at 218 (“Empirical statistics are not available to show that the inhabitants of states which follow the exclusionary rule suffer less from lawless searches and seizures than do those of states which admit evidence unlawfully obtained. Since as a practical matter it is never easy to prove a negative, it is hardly likely that conclusive factual data could ever be assembled. For much the same reason, it cannot positively be demonstrated that enforcement of the criminal law is either more or less effective under either rule.”).

⁶¹ “[W]e hold that evidence obtained by state officers during a search which, if conducted by federal officers, would have violated the defendant’s immunity from unreasonable searches and seizures under the Fourth Amendment is inadmissible over the defendant’s timely objection in a federal criminal trial.” *Id.* at 223.

For example, in *United States v. Calandra*,⁶² the Court characterized the exclusionary rule by noting that “the rule is a judicially created remedy designed to safeguard Fourth Amendment rights generally through its deterrent effect, rather than a personal constitutional right of the party aggrieved,”⁶³ and went on to hold that the exclusionary rule did not apply to grand jury proceedings. To justify this, the Court noted that despite the lack of evidence regarding the deterrent effect of the rule,

[a]ny incremental deterrent effect which might be achieved by extending the rule to grand jury proceedings is uncertain at best. Whatever deterrence of police misconduct may result from the exclusion of illegally seized evidence from criminal trials, it is unrealistic to assume that application of the rule to grand jury proceedings would significantly further that goal.⁶⁴

In other words, the Court acknowledged that it had no evidence of any deterrent effect of the exclusionary rule, but selectively used this lack of evidence to support its removal of grand jury proceedings from the purview of the rule.

The Court applied similar logic in *Stone v. Powell*,⁶⁵ when it declined to extend federal habeas relief to cover alleged violations of the exclusionary rule. First, the Court established that deterring police misconduct is the primary purpose of the exclusionary rule, stating that “[d]espite the absence of supportive empirical evidence, we have assumed that the immediate effect of exclusion will be to discourage law enforcement officials from violating the Fourth Amendment by removing the incentive to disregard it.”⁶⁶ Next, the Court justified its holding that violations of the rule should not be subject to habeas review by noting that “[t]he view that the deterrence of Fourth Amendment violations would be furthered rests on the dubious assumption that law enforcement authorities would fear that federal habeas review might reveal flaws in a search or seizure that went undetected at trial and on appeal.”⁶⁷ Again, the Court assumes the unsupported constitutional fact that the exclusionary rule acts as a government deterrent, and then selectively disregards a portion of its assumption to justify its result.

Perhaps most significantly, the Court used its technique of selectively undermining the empirical underpinnings of the exclusionary rule in *United States v. Leon*,⁶⁸ when it created a good faith exception to the rule. The Court cited *Calandra* for the proposition that the purpose of the rule is deterrence rather than preservation of an individual constitutional right. It then supported its creation of the good faith exception by stating “we discern no basis, and are offered none, for believing that exclusion of evidence seized pursuant to a warrant will have a significant deterrent effect on the issuing judge or

⁶² 414 U.S. 338 (1974).

⁶³ *Id.* at 348.

⁶⁴ *Id.* at 351.

⁶⁵ 428 U.S. 465 (1976).

⁶⁶ *Id.* at 492.

⁶⁷ *Id.* at 493.

⁶⁸ 468 U.S. 897 (1984).

magistrate.”⁶⁹ This is a strange justification, in light of the Court’s own acknowledgment of a complete lack of evidence regarding the deterrent effect of *any* use of the exclusionary rule.

As shown below in Figure 2, the evolution of the Court’s treatment of the exclusionary rule from its origins in *Weeks* to the Court’s creation of a good faith exception to the rule in *Leon* illustrates the Court’s willingness not only to rely on unsupported assertions of constitutional fact, but also, in a disturbing lack of candor, to selectively “chip away” at its own unsupported assertions to arrive at a desired result. Additional cases in which the Court used a similar strategy include *Walder v. United States*,⁷⁰ and *United States v. Janis*.⁷¹ In general, all of these cases show the Court’s willingness to invoke empirical considerations inappropriately, on the one hand making unsupported assertions of fact, and on the other hand using those unsupported assertions as “straw men” to be selectively knocked down.

B. Sixth Amendment Criminal Jury Size Cases and the Tensions Between Legal and Scientific Principles

In 1968, the Supreme Court extended the Sixth Amendment right to a jury trial to include criminal defendants in state court.⁷² However, whereas the size of a federal criminal jury had been fixed at twelve members by both long tradition⁷³ and case law,⁷⁴ the required size of state juries was initially unclear. In a line of cases extending through the 1970’s, the Court addressed the question of the constitutionally minimal size of a jury, ostensibly basing its decisions on the functional and empirically verifiable issue of verdict reliability. However, the Court’s interpretation of the available empirical data was questionable from the beginning, and illustrates many of the difficulties the Court faces when attempting to support its holdings with empirical data. In particular, the disparate flawed approaches taken by the Court in its treatment of data relevant to jury size include: 1) interpreting empirical data in an unscientific manner to “justify” a holding; 2) prematurely accepting evidence as conclusive, due to the immediacy of the Court’s need for empirical support; and 3) allowing stare decisis to act as a barrier to the acceptance of the new

⁶⁹ *Id.* at 916.

⁷⁰ 347 U.S. 62 (1954) (holding that unlawfully obtained evidence can be used for impeachment purposes at trial).

⁷¹ 428 U.S. 433 (1976) (holding that evidence attained by a state in violation of the Fourth Amendment can be used to prosecute a defendant for a federal civil offense).

⁷² *Duncan v. Louisiana*, 391 U.S. 145 (1968).

⁷³ According to one common view, the English tradition of twelve-member juries dates back to the reign of Henry II. *See, e.g.*, James B. Thayer, *The Jury and Its Development*, 5 HARV. L. REV. 295 (1892) (citations omitted) (“It seems to have been the recognitions under Henry II. that established twelve as the usual number . . .”).

⁷⁴ *See, e.g.*, *Thompson v. Utah*, 170 U.S. 343, 349 (1898) (holding that a jury referred to by the Sixth Amendment was a jury “constituted, as it was at common law, of twelve persons, neither more nor less.”).

knowledge. Each of these flawed approaches is the result of tensions that naturally exist between legal and scientific principles.

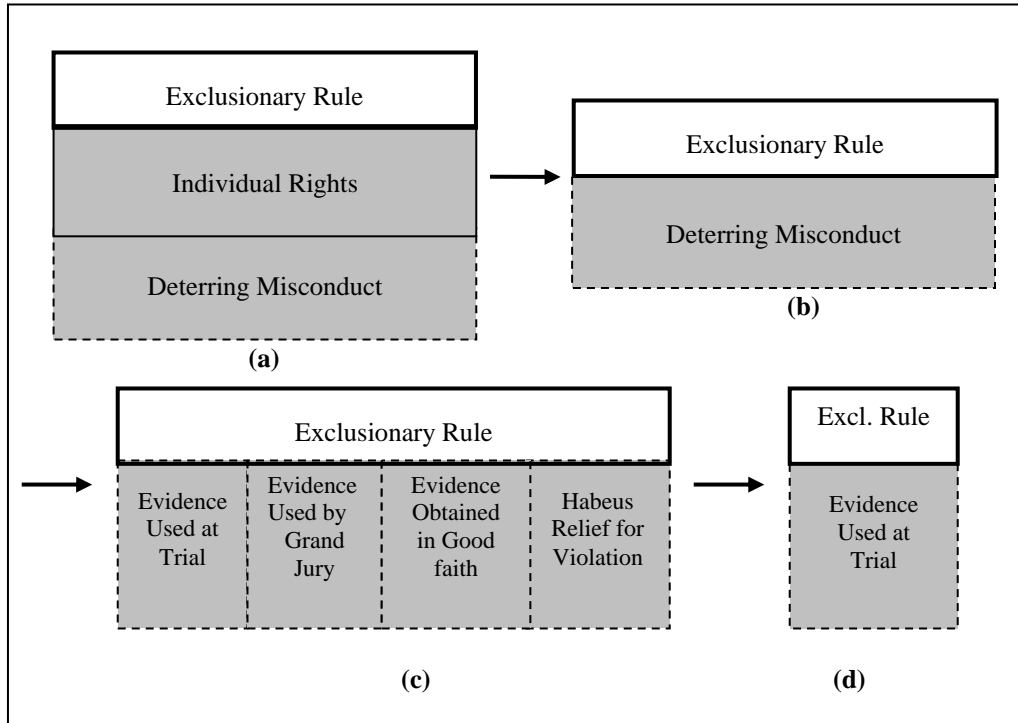


Figure 2

*The evolution of the Supreme Court's exclusionary rule jurisprudence. In (a), the Warren Court held that the dual rationales of individual rights and deterrence of government misconduct supported broad application of the rule. In (b), the post-Warren Court came to rely primarily on deterrence as support for the rule, despite a lack of empirical evidence for this rationale. In (c)–(d), the lack of empirical support for deterrence allowed the Court to treat the deterrent value of various types of improperly obtained evidence as “straw men,” which the Court proceeded to selectively knock down in *Calandra*, *Leon*, and *Stone v. Powell* to limit the scope of the rule.*

1. Anecdotal Versus Experimental Evidence

The Court began its consideration of jury size with *Williams v. Florida*⁷⁵ in 1970. At the time *Williams* was decided, only sparse anecdotal information was available regarding the effects of jury size on trial outcomes, but in an act revealing a startling lack of scientific sophistication, the Court cited this

⁷⁵ 399 U.S. 78 (1970).

information as empirically supportive of its position that twelve-member juries were not required in state criminal trials. Specifically, in finding a six-person jury sufficient, Justice White stated:

[i]t might be suggested that the 12-man jury gives a defendant a greater advantage since he has more ‘chances’ of finding a juror who will insist on acquittal and thus prevent conviction. But the advantage might just as easily belong to the State, which also needs only one juror out of twelve insisting on guilt to prevent acquittal. What few experiments have occurred—usually in the civil area—indicate that there is no discernible difference between the results reached by the two different-sized juries. In short, neither currently available evidence nor theory suggests that the 12-man jury is necessarily more advantageous to the defendant than a jury composed of fewer members.⁷⁶

This statement is illuminating for a number of reasons. In the first two sentences, the Court indulged in simplified theoretical speculation about a complicated question that inherently should be addressed empirically. In the third sentence, the Court characterized as “experiments” what can rightfully be termed only collections of anecdotes.⁷⁷ Finally, in the fourth sentence, the Court combined its theoretical speculation with its misguided reliance on the existing “experiments” to arrive at an unsupported conclusion.

Although from a scientific perspective, the Court’s mischaracterization of anecdotal information as experimental evidence in *Williams* seems egregious, it can be explained, at least in part, by a natural and fundamental difference between science and the law. Science seeks universal principles to explain a wide variety of observable phenomena, and such principles are separable from the particular perceptions of a given observer. Thus, an essential feature of a scientific experiment is its repeatability, and admissible scientific evidence is therefore by its nature entirely non-anecdotal. On the other hand, in a typical legal case, the facts are particularized and relate to past events that may be completely unrepeatable, so that large portions of the evidence must be entirely anecdotal, i.e. based on the perceptions of individual witnesses. In other words, evidence that would be absolutely valueless in a scientific inquiry is routinely admissible in court, and often determines the outcome of the case at issue.

Note, however, that the necessarily anecdotal nature of legal evidence applies only to evidence of particularized legal facts, and not to evidence of statistical facts. In *Williams*, the Court made the fundamental error of admitting anecdotal evidence to support an assertion of constitutional statistical fact. The Court’s obviously flawed reasoning in *Williams* led to a degree of outrage in the social science community, with one author declaring that “[t]he quality of

⁷⁶ *Id.* at 101–02 (citations omitted).

⁷⁷ In footnote 48 of the *Williams* opinion, the Court cited six “experiments” as support for its assertion, all of which have subsequently been extensively analyzed and found to be entirely anecdotal. *Id.* at 78 n.48. See, e.g., Robert H. Miller, *Six of One is Not a Dozen of the Other: A Reexamination of Williams v. Florida and the Size of State Criminal Juries*, 146 U. PA. L. REV. 621, 652 (1998) (noting that the Court’s evaluation of the experiments “was deeply flawed,” describing the anecdotal nature of the experiments, and citing numerous analyses supporting these conclusions).

social science scholarship displayed in those decisions would not win a passing grade in a high school psychology class.”⁷⁸ The Court attempted to correct its mistake a few years later when it next considered jury size, only to stumble into another pitfall stemming from tensions between the scientific and legal realms.

2. *Resolution Versus Reliability*

In 1973, the Court revealed another inherent conflict between legal and scientific evidence—this one related to timing—when it extended *Williams* by declaring the constitutionality of six-person juries in state civil trials. In its opinion in *Colgrove v. Battin*,⁷⁹ the Court noted that “[s]ince [*Williams*], much has been written about the six-member jury, but nothing that persuades us to depart from the conclusion reached in *Williams*.”⁸⁰ In a related footnote, the Court cited no less than 18 publications post-dating *Williams*, only four of which, however, represent the results of empirical studies.⁸¹ The Court characterized these studies as highly supportive of the notion that six-person juries are sufficient, stating that “four very recent studies have provided *convincing empirical evidence* of the correctness of the *Williams* conclusion that ‘there is no discernible difference between the results reached by the two different-sized juries.’”⁸² Unfortunately, each of these four studies is now widely viewed as severely methodologically flawed, rendering them scientifically highly questionable, if not entirely unsound.⁸³

As with the Court’s misguided reliance on anecdotal evidence in *Williams*, the Court’s reliance in *Colgrove* on what are now regarded as flawed studies may be understood in terms of an inherent distinction between the legal and scientific realms. In science, reliability of a result is considered to be of paramount importance, and a scientific hypothesis will typically be tested repeatedly, by many different groups of scientists and over a period of years, before it gains broad acceptance. Furthermore, the analysis performed by these groups will be checked and rechecked along the way, in an effort to avoid mistaken assertions of scientific fact. Aside from a small degree of professional pressure, there is little motivation for scientists to make premature and unverified claims, and in fact any mistakes along these lines are likely to affect

⁷⁸ Michael J. Saks, *Ignorance of Science Is No Excuse*, TRIAL, Nov.–Dec. 1974, at 18.

⁷⁹ 413 U.S. 149 (1973).

⁸⁰ *Id.* at 158-59.

⁸¹ *Id.* at 159 n.15. The results of the four studies are provided by: Note, *Six-Member and Twelve-Member Juries: An Empirical Study of Trial Results*, 6 U. MICH. J. L. REFORM 671 (1973); Note, *An Empirical Study of Six- and Twelve-Member Jury Decision-Making Processes*, 6 U. MICH. J. L. REFORM 712 (1973); Gordon Bermant & Rob Coppock, *Outcomes of Six- and Twelve-Member Jury Trials: An Analysis of 128 Civil Cases in the State of Washington*, 48 WASH. L. REV. 593 (1973); INST. OF JUDICIAL ADMIN., A COMPARISON OF SIX- AND TWELVE-MEMBER CIVIL JURIES IN NEW JERSEY SUPERIOR AND COUNTY COURTS (1972). The three journal articles are reproduced in JURY SIZE: ARTICLES AND BIBLIOGRAPHY FROM THE LITERATURE OF LAW AND THE SOCIAL AND BEHAVIORAL SCIENCES, (J. Myron Jacobstein & Roy. M. Mersky eds., 1998).

⁸² *Colgrove*, 413 U.S. at 159 n.15 (emphasis added).

⁸³ For a detailed discussion of the flawed methodology of each study, see, e.g., Miller, *supra* note 77, at 657–61, and references cited therein.

the career of the false claimant severely.⁸⁴ In law, however, immediate resolution of a controversy has inherent value that is independent of the “correctness” of the judgment, because a primary purpose of our legal system is to resolve controversies quickly. The tension between the scientific value of reliability and the legal emphasis on rapid resolution helps to explain the Court’s premature acceptance of the studies cited in *Colgrove*.⁸⁵

One significant indication of the lack of reliability of the *Colgrove* studies is that none of them were published in peer-reviewed journals, so that the methodology used in those studies had not been objectively independently evaluated.⁸⁶ This may have been due to a particular desire on the part of the authors to publish their studies as quickly as possible in the aftermath of *Williams*, or it may simply have been that non-peer-reviewed legal journals were the most natural publication forum available to those authors, but in any case the lack of peer review undoubtedly allowed publication of the studies without sufficient verification of the results they claimed to indicate.⁸⁷ The Court’s reliance on such studies in *Colgrove* indicates both a degree of scientific naiveté, and a strong bias towards quick legal resolution at the expense of scientific reliability.

Williams and *Colgrove* generated a large number of methodologically sound scholarly studies related to jury size, the essential consensus of which

⁸⁴ For example, after premature and unverifiable claims of cold fusion by Stanley Pons and Martin Fleischmann were discredited, both Pons and Fleischmann were driven from their academic positions. *See, e.g.*, Wikipedia, Stanley Pons, http://en.wikipedia.org/wiki/Stanley_Pons (last visited Apr. 11, 2006) (noting that both Pons and Fleischmann moved to France and accepted jobs for the Toyota Corporation after their cold fusion work was discredited). In a surprising twist, in recent years cold fusion has shown signs of regaining scientific credibility, although it is still controversial, and the original announcement made by Pons and Fleischmann remains the subject of scientific ridicule. *See, e.g.*, Aaron Galonsky et al., Letters, *Tabletop Fusion Revisited*, 297 SCIENCE 1645 (2002) (describing a tabletop experiment that produces anomalous heat, and suggesting that the heat may be produced by a form of nuclear fusion occurring at room temperature).

⁸⁵ An interesting analogy can be made between the law and applied (as opposed to empirical or theoretical) science, since both the law and applied science depend to some degree on underlying scientific findings, and both put a high premium on timely results. However, in the case of applied science, the value of speed stems from the societal gains resulting from advancing technology, and these gains do not purport to rely on the correctness of the principles underlying the applied techniques employed. In the legal realm, on the other hand, the value of speed stems from a preference for resolving cases, and carries with it a precedential effect that purports to rely on the validity of the evidence used to support the holding.

⁸⁶ *See supra* note 81. Three of the studies were published in law review journals, and the fourth was published by the Institute for Judicial Administration, a private organization founded by Arthur Vanderbilt and affiliated with the New York University School of Law. NYU School of Law, Institute of Judicial Administration, <http://www.law.nyu.edu/institutes/judicial/> (last visited Dec. 18, 2005). None of these publications were peer reviewed.

⁸⁷ The lack of peer review for articles published in legal journals may be viewed as another indication that reliability is less important in the realm of law than in the realm of science.

was that smaller juries are more likely to wrongfully convict than larger ones.⁸⁸ The Court's treatment of these studies illustrates yet another basic divide between science and jurisprudence.

3. *Revolution Versus Finality*

Five years after *Colgrove*, the Court had another occasion to review the question of jury size, but despite the availability of a large amount of reliable new empirical data related to the effects of jury size, the Court failed to act in accordance with the new information. In a fractured opinion in *Ballew v. Georgia*,⁸⁹ Justice Blackmun acknowledged the relevance of the new, post-*Williams* studies, stating “[w]e have considered them carefully because they provide the only basis, besides judicial hunch, for a decision about whether smaller and smaller juries will be able to fulfill the purpose and functions of the Sixth Amendment.”⁹⁰ After careful consideration of the post-*Williams* studies, the Court did in fact draw a number of correct conclusions based on the results of the studies, all of which indicated that twelve-person juries are more reliable than six-person juries. Remarkably, however, the Court failed to render a decision based on these conclusions.

Specifically, Blackmun's opinion in *Ballew* drew the following five conclusions:

First, recent empirical data suggest that progressively smaller juries are less likely to foster effective group deliberation. . . . The smaller the group, the less likely are members to make critical contributions necessary for the solution of a given problem. . . .⁹¹

Second, the data now raise doubts about the accuracy of the results achieved by smaller and smaller panels. Statistical studies suggest that the risk of convicting an innocent person (Type I error) rises as the size of the jury diminishes. . . . Another doubt about progressively smaller juries arises from the increasing inconsistency that results from the decreases. . . . 12-person groups reached correct verdicts 83% of the time; 6-person panels reached correct verdicts 69% of the time. . . .⁹²

Third, the data suggest that the verdicts of jury deliberation in criminal cases will vary as juries become smaller, and that the variance amounts to an imbalance to the detriment of one side, the defense. . . . [A] person in the minority will adhere to his position more frequently when he has at least one other person supporting his argument. . . .⁹³

⁸⁸ See *infra* note 92.

⁸⁹ 435 U.S. 223 (1978).

⁹⁰ *Id.* at 232 n.10.

⁹¹ *Id.* at 232–33.

⁹² *Id.* at 234–35. Three studies were cited to support the proposition that twelve-person groups reach accurate verdicts more often than six-person groups; in the text I have quoted the findings of only one of the three. A second study of former jurors found that twelve-person groups reach accurate verdicts 71% of the time as compared to 57% for six-person groups, and a third study found a significant difference in conviction propensity variation between groups of twelve-person juries and groups of six-person juries. *Id.* at 235.

⁹³ *Id.* at 236.

Fourth, what has just been said about the . . . decrease in size foretells problems not only for jury decisionmaking, but also for the representation of minority groups in the community. . . . [T]he opportunity for meaningful and appropriate representation does decrease with the size of the panels. . . .⁹⁴

Fifth, several authors have identified in jury research methodological problems tending to mask differences in the operation of smaller and larger juries. . . . Disparities . . . appear in only small percentages. Nationwide, however, these small percentages will represent a large number of cases. And it is with respect to those cases that the jury trial right has its greatest value.⁹⁵

As a result of its analysis of the new studies on the effects of jury size, a plurality of the *Ballew* Court drew five conclusions, each of which indicates that twelve-person juries more appropriately meet the goals of the Sixth Amendment than six-person juries. Yet incredibly, rather than overrule *Williams*, the Court stated that

[w]hile we adhere to, and reaffirm our holding in *Williams v. Florida*, these studies, most of which have been made since *Williams* was decided in 1970, lead us to conclude that the purpose and functioning of the jury in a criminal trial is seriously impaired, and to a constitutional degree, by a reduction in size to below six members.⁹⁶

Thus, the Court affirmed *Williams* despite its own unavoidable conclusion that six-person juries are far less reliable than twelve-person juries, and held only that criminal juries having *five* or fewer members are unconstitutionally small.

The Court's failure to render a decision in accordance with its own conclusions of statistical fact in *Colgrove* can be understood partially in light of the tension between the scientific method and stare decisis. In the world of modern science, progress is measured by the frequency with which preexisting ideas are overturned, and thus doctrinal changes, and even revolutionary paradigm shifts, are welcome. This is an essential feature of the scientific method, under which hypotheses are presented, tested, and then either adopted or discarded based on the results of the tests. In contrast, the law values finality, and legal principles may be upheld for many years on the basis of stare decisis. There is no provision in our justice system for legal hypotheses, and thus a legal proposition, once articulated, typically will remain in place until overruled by a jurisprudential shift, rather than an empirical one. This preference for finality in the law, which can be termed "legal inertia," gives rise to an inherent conflict when empirical data suggests that a law should be overruled.

4. *Postscript*

Finally, a year after *Colgrove*, the Court again reaffirmed *Williams* while holding that conviction by a six-person jury in a state criminal trial must be

⁹⁴ *Id.* at 236–37.

⁹⁵ *Id.* at 237–38.

⁹⁶ *Id.* at 239.

unanimous. In *Burch v. Louisiana*,⁹⁷ a majority opinion of the Court, authored by Justice Rehnquist, concluded that although “the question presented is a ‘close’ one . . . conviction by a nonunanimous six-member jury in a state criminal trial for a nonpetty offense deprives an accused of his constitutional right to trial by jury.”⁹⁸ Notably, the opinion makes no mention of empirical studies of the type so heavily relied upon in *Williams*, *Colgrove*, and *Ballew*. Rather, the opinion simply asserts that “the additional authorization of nonunanimous verdicts by such [six-person] juries sufficiently threatens the constitutional principles that led to establishment of the size threshold that any countervailing interest of the State should yield.”⁹⁹ In other words, the Court decided *Burch* based on constitutional principles, rather than constitutional statistical facts. The Court’s retreat from consideration of empirical data in its jury size jurisprudence indicates an implicit recognition of the difficulties of merging scientific and legal principles,¹⁰⁰ as well as an unwillingness by at least some members of the Court to meet those challenges.

IV. RESOLVING THE TENSIONS INHERENT IN CONSTITUTIONAL FACT FINDING

As the foregoing sections have illustrated, due to the conflicting goals of science and the law, the Supreme Court’s treatment of empirical data to date has been inconsistent and unscientific. Whereas science seeks reliability and consistency by demanding repeated and non-anecdotal evidence, the law by its nature requires some use of anecdotal evidence, and seeks consistency through the finality of its judgments. In other words, although both science and law provide checks on uncontrolled doctrinal changes, the temporal location of the primary control mechanism is different in each realm. As depicted below in Figure 3, science places its resistance to rapid changes temporally “up front,” by placing a high value on objectively verifiable experimentation, whereas the law places its resistance temporally “in back,” by placing a high value on stare decisis. These goals conflict when the Court attempts to integrate scientific

⁹⁷ 441 U.S. 130 (1979).

⁹⁸ *Id.* at 134.

⁹⁹ *Id.* at 139.

¹⁰⁰ After *Ballew*, the Court did not completely eliminate consideration of statistical studies from its jurisprudence, but appears to have limited its use of empirical studies to cases in which the evidence is particularly unequivocal. *See, e.g.*, *Batson v. Kentucky*, 476 U.S. 79, 104 (1986) (citing as support for the Court’s decision to disallow prosecutors from peremptorily challenging prospective jurors solely on the basis of race a study showing that “[i]n 100 felony trials in Dallas County in 1983–1984, prosecutors peremptorily struck 405 out of 467 eligible black jurors; the chance of a qualified black sitting on a jury was 1 in 10, compared to 1 in 2 for a white.”); *McCleskey v. Kemp*, 481 U.S. 279, 293 (1987) (“statistical proof normally must present a ‘stark’ pattern to be accepted as the sole proof of discriminatory intent under the Constitution”) (internal citations omitted). However, it should be noted that the Court in *McCleskey* found insufficient proof of discriminatory intent, although the primary study cited (the Baldus study) indicated that “defendants charged with killing white victims in Georgia are 4.3 times as likely to be sentenced to death as defendants charged with killing blacks.” *Id.* at 321 (Brennan, J., dissenting).

evidence into its jurisprudence. The result is often that the Court ignores scientific principles in favor of rapid resolution and finality.

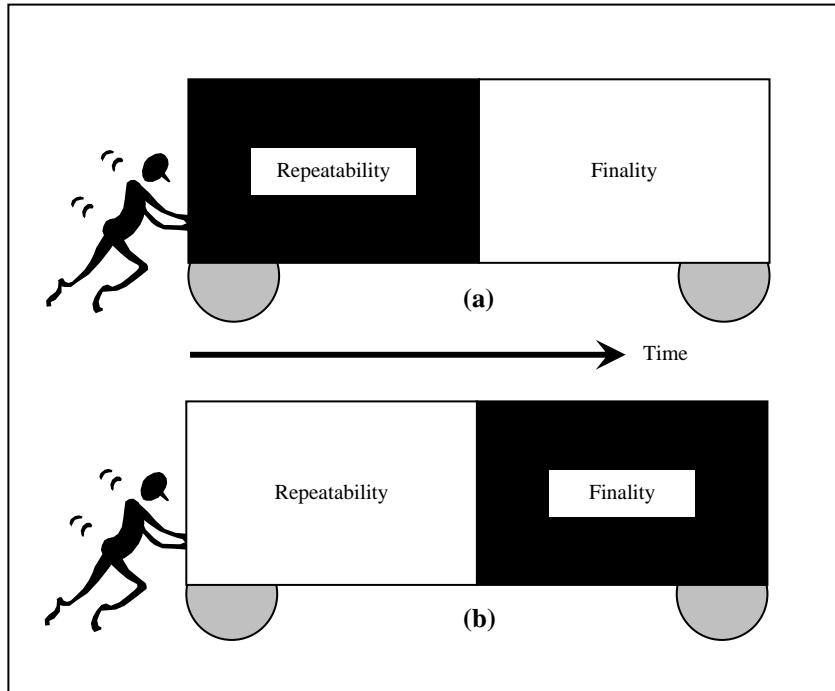


Figure 3

Sources of inertia in science and the law. In both realms, rapid evolution is checked by an inherent mechanism, but the mechanism is different in each case. In (a), the primary source of scientific inertia, indicated in black, is a demand for repeatability, forcing scientists to carefully verify their claims before advancing those claims as correct. In (b), the primary source of legal inertia, again indicated in black, is a preference for finality, so that stare decisis often bars change. Thus as the figures indicate, scientific inertia is “front-loaded,” whereas legal inertia is “rear-loaded.”

For the Court to resolve the inconsistencies in its treatment of empirical evidence, it should first distinguish those cases in which its holdings are driven by structural or doctrinal considerations from those in which the Court believes that empirical data must play a role. As described above,¹⁰¹ cases in which empirical evidence may be inherently irrelevant include separation of powers cases, federalism cases, and many equal protection cases in which the level of scrutiny is either so high or so low as to make empirical considerations moot. A first step toward a more defensible use of empirical evidence should be the

¹⁰¹ See *supra* Part II.

Court's acknowledgment that its use of such data is only appropriate in certain cases, and an enumeration of what types of cases might be suitable for empirical analysis.

Even in areas where the Court believes empirical data is relevant, to incorporate empirical evidence into its jurisprudence in a consistent and principled manner, the Court first must recognize the sources of its past inconsistency, and be willing to change its behavior. As I have argued in this Paper, one such source of inconsistency is an inherent and fundamental difference between the goals of science and the goals of law, but this is not the only reason for the Court's unprincipled treatment of scientific studies. As the Court's exclusionary rule holdings illustrate (*see infra* Part III.A), the Court has shown a willingness to invoke empirical considerations in a disingenuous manner, to mask an ulterior motive. This lack of judicial candor is incompatible with scientific principles, and precludes the proper use of empirical evidence. The inherent tensions between science and law can only be resolved in cases where the Court has no hidden agenda, and is willing to candidly accept the implications of the available empirical evidence. In other words, to act in accordance with scientific principles, the Court must make a threshold determination of the relevant scope of empirical data, articulate those areas of legal relevance, and then consistently and candidly consider available data in cases dealing with those areas of law.

Assuming the Court actually desires to candidly consider empirical data in certain areas of law, it then must institute procedures for considering the data in a scientific manner. Members of the Court are not scientists, and history has shown the ease with which scientific evidence can be mischaracterized or misinterpreted in the hands of non-scientists. The Court's flawed use of empirical studies in *Williams* and *Colgrove* illustrates the Court's own propensity for errors of this kind. To safeguard against mistakes, at a minimum the Court should only consider peer-reviewed scientific studies, and should avoid considering studies published in law review journals. While law review articles are an appropriate format for theoretical legal scholarship, the lack of peer review in legal journals makes empirical studies published in those journals less reliable than those published in peer-reviewed, social science journals. As an even more effective measure, the Court should have access to a nonpartisan scientific advisory committee, the purpose of which would be to evaluate the reliability of empirical studies under consideration by the Court, and possibly to help interpret the results of those studies. This could be a newly formed proprietary committee, perhaps formed under the auspices of the National Science Foundation (NSF),¹⁰² or the Court could be given access to an existing scientific advisory agency, such as the United States Office of Science and Technology Policy.¹⁰³

¹⁰² NSF was created by Congress in 1950 primarily "to promote the progress of science." U.S. Nat'l Science Found., About the National Science Foundation, <http://www.nsf.gov/about/> (last visited Apr. 11, 2006).

¹⁰³ However, separation of powers might be a concern with this particular agency, since the agency was established by Congress in 1976 "to advise the President and others within

Finally, in areas where the Court wishes to allow empirical findings to have a genuine impact on the law, the Court either must delay consideration of the data until it demonstrates unequivocal statistical facts, or it must relax *stare decisis* to allow the scientific method to clarify prior results, and in some cases overturn those results. Since in many instances, relevant data will not become available until after the Court makes a ruling,¹⁰⁴ the second option—a selective relaxation of *stare decisis*—is theoretically preferable. The Court could accomplish this, for example, by creating a new class of provisional rulings, with the understanding that final resolution of the issue at hand would be revisited upon a showing that the issue had been sufficiently empirically studied. Although such a scheme poses certain practical problems related to how a relatively rapidly evolving standard of law would be applied,¹⁰⁵ tackling these problems is preferable to the Court's continued sporadic, inconsistent, and unprincipled treatment of empirical data.

the Executive Office of the President on the effects of science and technology on domestic and international affairs." Office of Science and Technology Policy, What We Do, http://www.ostp.gov/html/_whatwedo.html (last visited April 11, 2006).

¹⁰⁴ For example, in *United States v. Ruiz*, the Court held that prosecutors need not turn over impeachment evidence to defense counsel prior to plea negotiations, stating "[i]t is particularly difficult to characterize such information as critical, given the random way in which it may, or may not, help a particular defendant." 536 U.S. 622, 623 (2002). The assertion that defense counsel's advance knowledge of evidence impeaching government witnesses has a "random" effect on pleas is both intuitively suspect and empirically testable. However, since the law prior to *Ruiz* required prosecutors to turn over the impeachment evidence, an empirical comparison only became possible after *Ruiz* was decided. See *Giglio v. United States*, 405 U.S. 150 (1972) (setting out the requirement).

¹⁰⁵ For instance, changes in the provisional law should be applied to cases on habeas review, which would require that such changes not be considered "new rules" for purposes of habeas consideration. See *Teague v. Lane*, 489 U.S. 288 (1989) (holding that changes in the law shall not be applied retroactively to cases on collateral review).