



UNC
SCHOOL OF LAW

NORTH CAROLINA LAW REVIEW

Volume 95 | Number 3

Article 3

3-1-2017

Tyson Foods and the Future of Statistical Adjudication

Robert G. Bone

Follow this and additional works at: <http://scholarship.law.unc.edu/nclr>



Part of the [Law Commons](#)

Recommended Citation

Robert G. Bone, *Tyson Foods and the Future of Statistical Adjudication*, 95 N.C. L. REV. 607 (2017).

Available at: <http://scholarship.law.unc.edu/nclr/vol95/iss3/3>

This Article is brought to you for free and open access by Carolina Law Scholarship Repository. It has been accepted for inclusion in North Carolina Law Review by an authorized editor of Carolina Law Scholarship Repository. For more information, please contact law_repository@unc.edu.

TYSON FOODS AND THE FUTURE OF STATISTICAL ADJUDICATION*

ROBERT G. BONE**

Statistical adjudication, the practice of using sampling and other statistical techniques to adjudicate large case aggregations, is highly controversial today. In all its forms, statistical adjudication decides cases on the basis of statistical extrapolation rather than case-specific facts. For example, a court adjudicating a large class action might try a random sample of cases, average the trial verdicts, and give the average to all the other cases in the aggregation. In Wal-Mart Stores, Inc. v. Dukes, the Supreme Court rejected a sampling proposal as inconsistent with the Rules Enabling Act, calling it “Trial by Formula.” In the wake of this decision, at least one commentator declared the death of statistical adjudication.

In an important decision last term, Tyson Foods, Inc. v. Bouaphakeo, the Court changed course and breathed new life into statistical adjudication. It upheld the use of sampling to establish liability and damages in a Fair Labor Standards Act case and indicated that sampling might be available in other cases as well. The Court’s opinion is far from clear, however, and offers little guidance to lower court judges trying to determine when and how to use the procedure in future cases.

This Article explores the impact of Tyson Foods on the future of statistical adjudication. Part I defines statistical adjudication and distinguishes it from statistical evidence. Part II shows that Tyson Foods is better understood as a case of statistical adjudication than simply a case of statistical evidence. Part III takes a closer look at the Court’s opinion in an effort to tease out factors and principles to guide the future use of statistical adjudication. Part

* © 2017 Robert G. Bone.

** G. Rollie White Professor of Law, University of Texas at Austin School of Law. I am especially grateful to my colleague Patrick Woolley for his very helpful comments and conversations. I also wish to thank Beth Burch, Jane Cohen, Ward Farnsworth, Jonah Gelbach, Alexandra Lahav, Larry Sager, Jay Tidmarsh, and Steve Vladek for reading earlier drafts and providing useful input, as well as participants in the University of Texas School of Law Faculty Colloquium. Finally, I am grateful to Michael Davis for his excellent research assistance.

IV explores reasons for the vague discomfort with the procedure, reasons that seem tied to nagging doubts about its legitimacy. Critics worry that statistical adjudication is too strange a fit with adjudication, too substantive to be legitimately implemented as procedure, and too mechanical to count as a proper form of adjudicative reasoning. Part IV argues that statistical adjudication is not as strange as it might seem, that its outcome effects do not make it too substantive, and that while it substitutes a mechanical decision algorithm for the usual reasoning process, it does so in a way that can be justified as legitimate. It is time that we recognize statistical adjudication for what it is: a useful procedural tool that, when carefully designed and selectively deployed, is capable of adjudicating large case aggregations fairly and efficiently.

INTRODUCTION	609
I. STATISTICAL ADJUDICATION DEFINED	611
II. <i>TYSON FOODS</i> : A CRITICAL SUMMARY	616
A. <i>The Background Facts</i>	616
B. <i>The Court's Holding</i>	620
C. <i>The Court's Reasoning</i>	622
1. Evidentiary Inference or Statistical Adjudication?	623
2. Inference of Liability and Damages?	627
3. Trial by Formula?	629
III. THE FUTURE OF SAMPLING AFTER <i>TYSON FOODS</i>	633
A. <i>The Court's Factors</i>	633
1. The Purpose of Sampling	634
2. The Underlying Cause of Action	639
3. Heterogeneity and Reliability	642
B. <i>Justifying Sampling's Outcome Effects</i>	643
C. <i>Other Factors</i>	649
IV. THE LEGITIMACY PROBLEM	654
A. <i>Normative Legitimacy Versus Perceived Legitimacy</i>	655
B. <i>Three Arguments from Normative Legitimacy</i>	655
1. Sampling Is Too Strange	657
a. <i>Sampling Is Not That Strange After All</i>	658
b. <i>The Value of Sensible Reform</i>	661
2. Sampling Is Too Substantive	661
3. Sampling Is Too Mechanical	665
CONCLUSION	671

INTRODUCTION

Statistical adjudication, the practice of using sampling and other statistical techniques to adjudicate large case aggregations, is highly controversial. In all its various forms, statistical adjudication produces decisions based not on the facts of the specific case but on statistical extrapolations from results in a sample of other cases.¹ For example, a court adjudicating a large class action might try a random sample of cases, average the sample verdicts, and either give the average to all the other cases in the aggregation or calculate an aggregate damages award for the class based on the average. When properly designed, statistical adjudication can produce a reasonably accurate aggregate liability figure, reduce the total private and public costs of litigation, and sometimes save enough in litigation costs to make plaintiffs better off than they would be with individual trials. Notwithstanding these benefits, courts have always had serious misgivings about the procedure. Only a few federal district courts have used it, and many of those courts have been reversed on appeal.² In its 2011 decision in *Wal-Mart Stores, Inc. v. Dukes*,³ the Supreme Court dismissed a proposed sampling procedure with the pejorative label “Trial by Formula” and held that it exceeded the limits of the Rules Enabling Act (“REA”).⁴ In the wake of this decision, at least one commentator declared the death of statistical adjudication.⁵

1. For examples of statistical adjudication, see *Hilao v. Estate of Marcos*, 103 F.3d 767, 782–84 (9th Cir. 1996); *Cimino v. Raymark Indus., Inc.*, 751 F. Supp. 649, 653 (E.D. Tex. 1990), *vacated in part*, 151 F.3d 297 (5th Cir. 1998).

2. See *Blue Cross & Blue Shield of N.J., Inc. v. Philip Morris, Inc.*, 178 F. Supp. 2d 198, 250–62 (E.D.N.Y. 2001), *rev'd in part*, 344 F.3d 211 (2d Cir. 2003), *rev'd sub nom.* *Empire Healthchoice, Inc. v. Philip Morris USA Inc.*, 393 F.3d 312 (2d Cir. 2004) (*per curiam*); *In re Estate of Marcos Human Rights Litig.*, 910 F. Supp. 1460, 1462 (D. Haw. 1995), *aff'd sub nom.* *Hilao v. Estate of Marcos*, 103 F.3d 767 (9th Cir. 1996); *Cimino*, 751 F. Supp. at 653; Jay Tidmarsh, *Resurrecting Trial by Statistics*, 99 MINN. L. REV. 1459, 1461 n.10 (2015) (reviewing cases in which sampling procedures have been rejected as well as cases that show unease about the use of sampling).

3. 564 U.S. 338 (2011).

4. See *id.* at 367.

5. Tidmarsh, *supra* note 2, at 1459 (noting that after *Wal-Mart* “[t]he notion that a court could try a representative sample of monetary claims and extrapolate the average result to the remainder of the cases was finished”). For other discussions of statistical adjudication published after *Wal-Mart*, see Hillel J. Bavli, *Aggregating for Accuracy: A Closer Look at Sampling and Accuracy in Class Action Litigation*, 14 LAW, PROBABILITY & RISK 67, 82–83 (2015); Alexandra D. Lahav, *The Case for “Trial by Formula,”* 90 TEX. L. REV. 571, 577–79 (2012). See generally Laurens Walker & John Monahan, *Sampling Damages*, 83 IOWA L. REV. 545 (1998) [hereinafter Walker & Monahan, *Sampling Damages*] (arguing that the use of random sampling for damages can facilitate mass tort

This death notice was premature. In an important decision last term, *Tyson Foods, Inc. v. Bouaphakeo*,⁶ the Supreme Court changed course and breathed new life into statistical adjudication. But the Court's opinion is far from clear. It treats the case as one involving statistical evidence and employee-specific inferences when it actually involves substituting statistical averages for employee-specific fact finding. This makes it more like a case of statistical adjudication than a case of statistical evidence. Moreover, while the Court contemplates the use of statistical adjudication more broadly, its opinion offers little explicit guidance to lower court judges who must determine when and how to use the procedure in future cases.

Thus, while statistical adjudication is alive, its prognosis is uncertain. But why? Why the discomfort with using statistical techniques to extrapolate from sample results to outcomes in related cases? The concern seems to be about more than just the overall balance of social costs and benefits or the impact on participation and substantive rights. When objections are framed in legal terms, they often invoke institutional limits on what courts can properly do. But what are those limits? And why does a seemingly sensible procedure for fairly and efficiently resolving large aggregations of related cases exceed those limits?

This Article explores the impact of *Tyson Foods* on the future of statistical adjudication. It builds on my earlier work analyzing the costs and benefits of statistical adjudication and extends that work by developing its implications in light of the *Tyson Foods* decision.⁷ The Article has two main goals. First, it seeks to fill the gap that the *Tyson Foods* Court left open and develop a sensible set of principles and factors to guide future use of statistical adjudication. Second, it probes the reasons for general discomfort with the procedure. Unless that discomfort is diagnosed and critically examined, statistical adjudication is unlikely to fare well, even in those cases where it otherwise makes sense on policy grounds.

The body of this Article is divided into four parts. Part I defines statistical adjudication and distinguishes it from statistical evidence.

litigation); Laurens Walker & John Monahan, Essay, *Sampling Liability*, 85 VA. L. REV. 329 (1999) (arguing that sampling for liability makes sense in a range of cases).

6. 136 S. Ct. 1036 (2016).

7. For this earlier work, see Robert G. Bone, *A Normative Evaluation of Actuarial Litigation*, 18 CONN. INS. L.J. 227, 231–33 (2011) [hereinafter Bone, *Actuarial Litigation*]; Robert G. Bone, *Statistical Adjudication: Rights, Justice, and Utility in a World of Process Scarcity*, 46 VAND. L. REV. 561, 566–69 (1993) [hereinafter Bone, *Statistical Adjudication*].

Part II provides background on *Tyson Foods* and examines the Court's opinion with care. It shows that *Tyson Foods* is better understood as a case of statistical adjudication than a case involving only statistical evidence. It also highlights an important aspect of the decision: the Court extends statistical adjudication beyond estimating damages to determining liability as well.

Part III argues that *Tyson Foods*, when properly understood, authorizes statistical adjudication well beyond the Fair Labor Standards Act ("FLSA") claim at issue in the case. Part III also develops a set of principles and factors to guide future use of the procedure. It does so by carefully examining what the Court says and does in light of the broader purposes of civil adjudication and the impact of sampling on outcome quality and participation rights.

Part IV explores reasons for the vague discomfort with statistical adjudication. This discomfort, I believe, is tied to nagging doubts about the legitimacy of the procedure. Critics worry that statistical adjudication is too strange a fit with adjudication, too substantive to be properly implemented as procedure, and too mechanical to count as a legitimate form of adjudicative reasoning. Part IV addresses each of these concerns in turn. It argues that statistical adjudication is not as strange as it might seem, that its outcome effects do not make it too substantive, and that while it substitutes a mechanical decision algorithm for the usual process of judicial reasoning, it does so in a way that can be justified as legitimate.

I. STATISTICAL ADJUDICATION DEFINED

There are many uses for statistics in civil adjudication.⁸ The substantive law sometimes explicitly incorporates a probabilistic element that invites statistical proof, such as the "likelihood of confusion" test for liability in trademark law.⁹ Sometimes the law includes an element that is best evaluated statistically. For example, in a Title VII disparate impact claim, the differential effects of the defendant's employment practices must be measured in statistical

8. For some examples, see Joseph B. Kadane, *Probability Sampling in Litigation*, 18 CONN. INS. L.J. 297, 299-303 (2011).

9. See 15 U.S.C. § 1125(a)(1)(A) (2012). Liability for trademark infringement depends on whether the defendant's use of a mark is likely to cause consumer confusion, and sampling consumer reaction is a common way to prove the requisite likelihood. See 4 J. THOMAS MCCARTHY, MCCARTHY ON TRADEMARKS AND UNFAIR COMPETITION §§ 23:1, 23:2.50 (4th ed. 2016).

terms.¹⁰ And sometimes the law points to counterfactual propositions that can only be proved by statistical methods. For example, statistical modeling is used in antitrust suits to determine damages when it is impossible to know directly what the counterfactual market free from the antitrust violation would have looked like.¹¹

In all these examples, the substantive issue itself is statistical or makes sense only in statistical terms. As a result, statistical evidence is the obvious—and often the only—way to prove the issue and generate a reasonably correct substantive result for each individual case. In a trademark suit, for example, likelihood of consumer confusion is statistical by definition, and statistical sampling of prospective consumers is the best way to determine it. Moreover, disparate impact in a Title VII suit is essentially a statistical concept calling for statistical proof. And in an antitrust suit, when the substantive law requires proof of a counterfactual scenario for which there can be no direct evidence, statistical modeling is an obvious way to proceed.

Statistical adjudication is different. It employs aggregate statistics to decide issues that are not intrinsically statistical and for which, in theory, there could be nonstatistical, case-specific evidence. Thus, statistical adjudication does not focus on the individual case or try to generate a correct substantive result for each case based on case-specific facts. Instead, it treats each case as an average case and substitutes the sample average (or other statistic) for a case-specific result based on individualized fact-finding.¹² For example, when a judge adjudicates a large aggregation of cases by taking a random sample and averaging the sample case verdicts, the judge does not use the sample average as evidence from which to infer specific causation, contributory fault, damages, and the like for each individual case. That would be using the sample average as statistical evidence. Rather, the judge applies the sample average in place of individual facts and generates an average outcome for each plaintiff or an aggregate award for the class as a whole.

10. See GEORGE RUTHERGLEN, *EMPLOYMENT DISCRIMINATION LAW: VISIONS OF EQUALITY IN THEORY AND DOCTRINE* 79–89 (3d ed. 2010).

11. See, e.g., *Zenith Radio Corp. v. Hazeltine Research, Inc.*, 395 U.S. 100, 116 & n.11 (1969) (allowing the computation of antitrust damages based on a counterfactual calculation that plaintiffs would have enjoyed 16% market share as opposed to their actual 3.2% share).

12. For more background on statistical adjudication, see generally Bone, *Statistical Adjudication*, *supra* note 7.

There are other ways to implement statistical adjudication. One can, for example, use the sample average as a presumption that shifts the burden to the defendant when the defendant has no evidence to offer in rebuttal. Indeed, in *Tyson Foods*, as we shall see, the Court assumed that the defendant could not rebut the presumption and, therefore, that the sample average controlled the outcome for all class members just as though the presumption had been conclusive.¹³

A particularly important feature distinguishing statistical adjudication from statistical evidence has to do with the reason why the sample average is allowed to affect or control outcomes. For statistical evidence, that reason is to achieve the best possible jury or judge determination of the relevant issue based on the facts of an individual case. Since the aim is to get as close to the right decision as possible, the focus is on the probative value of the sample average compared to other available evidence.

The focus of statistical adjudication is different. The primary reason to use the sample average in statistical adjudication is not to get the right result on the facts of each individual case. Instead, it is to get a good enough result for all cases so that the substantive law achieves its deterrence goals and litigants are treated fairly and justly in relation to one another. As we shall see, the use of the sample average in *Tyson Foods* overcomes proof problems caused by the defendant's failure to keep adequate records and does so in a way that supports collective adjudication of the small claims involved and enables private enforcement of the FLSA.¹⁴

The distinction between statistical adjudication and statistical evidence is not always perfectly clear or precise. For example, an evidentiary presumption designed to force disclosure of information aids in enforcing the substantive law. Thus, one might use a sample average to support a presumption that shifts the burden to the defendant to force the defendant to disclose private information. This would be use as statistical evidence if the defendant is likely to have the information and its disclosure would improve decisional accuracy on the facts of the individual case. In *Tyson Foods*, however, the focus is not on ferreting out private information or getting as close as possible to an accurate estimate of overtime in each individual case. Instead, the focus is on generating a rough average estimate of

13. See *infra* notes 58–60, 87–88 and accompanying text.

14. See *infra* Section III.A.1.

overtime for all employees in order to overcome impediments to suit and optimally serve the policies of the substantive law.¹⁵

Stated simply, it is one thing to use the sample average because the best interpretation of the substantive law authorizes its use and it has probative value, or even because it supports a presumption that is designed to ferret out case-specific evidence relevant to deciding an individual issue when there is a reasonable expectation that the defendant has evidence to offer (so most cases will be decided in the ordinary way). It is quite a different thing to use the sample average, when the substantive law requires an individualized determination, on the ground that there are case-specific procedural or evidentiary problems with generating an individualized decision and enabling enforcement of the substantive law. The former is use as statistical evidence; the latter is use for statistical adjudication.

No matter what form it takes, statistical adjudication involves sampling. There are different ways to sample cases from an aggregation. One might draw a single sample or one might divide the aggregation into subgroups and sample each subgroup separately.¹⁶ The latter approach, known as stratified sampling, is more costly, but it is also useful in coping with greater population heterogeneity.¹⁷

The simplest way to employ statistical adjudication is to try a random sample of cases and average the sample results.¹⁸ To illustrate, suppose there are 100 aggregated cases. Rather than try each case individually, the judge randomly samples ten cases and tries each sampled case to a jury. Suppose that the jury finds for the defendant in two of the ten cases, and with respect to the other eight, it returns a verdict of \$100,000 in two of those eight cases, \$500,000 in four, and \$1 million in two. The average of the sample results is

15. See *infra* Section II.B.

16. For example, the trial judge in *Cimino v. Raymark Industries, Inc.*, 751 F. Supp. 649 (E.D. Tex. 1990), *vacated in part*, 151 F.3d 297 (5th Cir. 1998), used a stratified sample, dividing the asbestos class into five disease categories and sampling from each separately. See *id.* at 653.

17. It is also possible, in theory at least, to construct a linear regression equation from the sample results and use it to generate outcomes that vary with different case characteristics. See Bone, *Statistical Adjudication*, *supra* note 7, at 584–87. Properly done, linear regression can produce results that are even more closely tailored to the particular facts of individual cases, but it is even more costly to implement than stratified sampling.

18. The trial judge in the *Cimino* case proceeded in this way although he let the plaintiffs in the sampled cases keep their actual verdicts and gave the average to all the other plaintiffs in the aggregation. *Cimino*, 751 F. Supp. at 653.

$[(2 \times 0) + (2 \times 100,000) + (4 \times 500,000) + (2 \times 1,000,000)] \div 10 = \$420,000$.¹⁹ The court then enters a judgment for an aggregate award of $100 \times 420,000 = \$42,000,000$ —or gives each plaintiff an average award of $\$420,000$.²⁰

In a variant on this approach, instead of trying each case in the sample, the judge appoints a special master to survey and depose the sampled plaintiffs.²¹ Based on the results of these surveys and depositions, the special master recommends average damages for the population as a whole or average damages for each subgroup of a stratified sample.²² The judge submits these averages to a jury, along with expert testimony about the statistical reliability of the sampling procedure and perhaps some testimony from the sampled plaintiffs.²³ The jury decides whether to accept, modify, or reject the sample average and then returns an aggregate verdict.²⁴

It is important to bear in mind that even when a jury alters the sample results, it does not do so in order to determine the facts of each individual case. The reason the jury hears evidence beyond the sampling study is to evaluate the reliability of the methodology and possibly make adjustments to account for overall characteristics of the aggregation. The jury cannot find individual case-specific facts because it has no evidence for individual cases beyond those in the sample, and as we shall see, it cannot draw reliable case-specific inferences unless the aggregation is unusually homogeneous.²⁵

In these examples, statistical adjudication is used to determine both liability and damages. However, it is also possible to prove liability in the usual case-specific way and use sampling only to

19. There is another way to do this that yields the same result. First, note that two verdicts for the defendant implies a 20% invalidity rate for the entire population of cases. So, one might first apply the 20% invalidity rate to yield eighty cases in which plaintiffs should receive verdicts. Focusing on just the eight sample cases with plaintiff verdicts, the average is $[(2 \times 100,000) + (4 \times 500,000) + (2 \times 1,000,000)] \div 8 = \$525,000$. As a result, each plaintiff receives $525,000 - (0.2 \times 525,000) = \$420,000$. This is the method the special master in *Hilao* appears to have used. See *Hilao v. Estate of Marcos*, 103 F.3d 767, 782–84 (9th Cir. 1996).

20. Or it might give the sampled cases their own individual awards and the rest of the cases the sample average.

21. The method described in this paragraph is roughly the method used in *Hilao*, 103 F.3d at 782–84, and also the method proposed by the Ninth Circuit for determining backpay awards in *Dukes*, see *Dukes v. Wal-Mart Stores, Inc.*, 603 F.3d 571, 627 (9th Cir. 2010), *rev'd*, 564 U.S. 338 (2011).

22. See *Hilao*, 103 F.3d at 783.

23. See *id.* at 784.

24. See *id.*

25. See *infra* text accompanying notes 158–60.

measure damages.²⁶ In addition, statistical adjudication can be used to decide only particular issues. As we shall see, the court in *Tyson Foods* used statistical adjudication to determine a key factual issue critical to both liability and damages—the time taken to don and doff specialized equipment—and it did so by admitting into evidence a sampling study and allowing the jury to modify the sample average.²⁷

Finally, judges dealing with large aggregations sometimes select a sample of cases for so-called “bellwether trials.”²⁸ The purpose of holding bellwether trials is to facilitate settlement by generating a common baseline of trial verdicts to help estimate a settlement value for the aggregation.²⁹ Although the bellwether trial procedure involves sampling, it is distinct from statistical adjudication. The goal of bellwether trials is to facilitate settlement; the goal of statistical adjudication is to render a final judgment binding on all cases in the aggregation.

A point of clarification is in order before proceeding. This Article frequently refers to “sampling” without also mentioning statistical adjudication. Virtually all of these references assume sampling is being used as part of a statistical adjudication procedure. Where this is not the case, it will be clear from the context.

II. *TYSON FOODS*: A CRITICAL SUMMARY

A. *The Background Facts*

Tyson Foods was a wages-and-hours case brought by employees of Tyson Foods, Inc.’s pork processing plant in Storm Lake, Iowa.³⁰ The lawsuit, filed in the United States District Court for the Northern District of Iowa, alleged claims under the FLSA and the Iowa Wage Payment Collection Law (“IWPCCL”) to recover damages for

26. See Walker & Monahan, *Sampling Damages*, *supra* note 5, at 549, 561–65.

27. See *infra* Section II.A.

28. See *In re Vioxx Prods. Liab. Litig.*, 574 F. Supp. 2d 606, 608 (E.D. La. 2008); Eldon E. Fallon, Jeremy T. Grabill & Robert Pitard Wynne, *Bellwether Trials in Multidistrict Litigation*, 82 TUL. L. REV. 2323, 2326–29 (2008); Alexandra D. Lahav, *Bellwether Trials*, 76 GEO. WASH. L. REV. 576, 577–78 (2008) (analyzing the merits of bellwether trials).

29. Lahav, *supra* note 28, at 577–78.

30. See *Tyson Foods, Inc. v. Bouaphakeo*, 136 S. Ct. 1036, 1042 (2016). For an interesting discussion of the parties’ litigation strategy choices and their impact, see generally Andrew J. Trask, *Litigation Matters: The Curious Case of Tyson Foods v. Bouaphakeo*, 2016 CATO SUP. CT. REV. 279.

uncompensated overtime pay.³¹ The plaintiffs sought to represent other similarly situated employees under a FLSA collective action for the FLSA claims and under a Rule 23(b)(3) class action for the IWPCl claims.³²

The dispute arose out of Tyson's method of compensating employees in the kill, cut, and retrim departments at the plant.³³ The work involves slaughtering and preparing meat for market, which requires the use of knives and other sharp implements, and as a result, employees wear protective gear to safeguard against injury.³⁴ The issue in the case had to do with Tyson's responsibility to pay for the time employees spent donning and doffing their protective gear.³⁵

Tyson's so-called "gang-time" system compensates employees for the time they actually spend at their workstations but not for time spent donning and doffing.³⁶ In 1998, Tyson began compensating for donning and doffing time, but rather than keep track of the minutes for each employee, the company used an average figure of four minutes per day, which it called "K-code" time.³⁷ K-code time was changed in 2007 so that only some employees received it, and those who did were credited with between four and eight minutes, depending on their equipment.³⁸ Thus, throughout the period covered by the lawsuit, Tyson paid all employees for gang-time and some employees for gang-time plus K-code time.

The plaintiffs alleged that for many Storm Lake employees the time spent donning and doffing protective equipment exceeded the K-code time and that, as a result, their actual compensable working

31. *Tyson Foods*, 136 S. Ct. at 1042. The Court makes clear that the IWPCl claim targets the same allegedly wrongful conduct as the FLSA claim. *Id.* ("This statute provides for recovery under state law when an employer fails to pay its employees 'all wages due,' which includes FLSA-mandated overtime" (quoting IOWA CODE ANN. § 91A.3 (West, Westlaw through 2016 Reg. Sess.))).

32. *See id.* The FLSA collective action is created by the FLSA statute, *see* 29 U.S.C. § 216 (2012), whereas the Rule 23(b)(3) class action is created by the Federal Rules of Civil Procedure, *see* FED. R. CIV. P. 23(b)(3). The FLSA collective action is an opt-in procedure; employees must affirmatively elect to join the action in order to benefit from it and be bound by it. *See* *Genesis Healthcare Corp. v. Symczyk*, 133 S. Ct. 1523, 1530 (2013). The Rule 23(b)(3) class action, by contrast, is an opt-out procedure; class members must affirmatively choose to exit the class if they wish not to be bound by the result. *See* FED. R. CIV. P. 23(c)(2)(B)(v).

33. *See Tyson Foods*, 136 S. Ct. at 1042.

34. *See id.*

35. *Id.* at 1041.

36. *Id.* at 1042.

37. *Id.*

38. *Id.*

time exceeded forty hours per week, entitling them to overtime pay.³⁹ Because Tyson kept no records of donning and doffing times, the plaintiffs had no basis for calculating individual overtime.⁴⁰ They relied instead on two studies: a statistical study by Dr. Kenneth Mericle that estimated average donning and doffing time for a sample of employees and a study by Dr. Liesl Fox that used the results of the Mericle study to calculate the overtime pay owed by Tyson.⁴¹

The Mericle study employed a statistical sampling procedure to estimate average donning and doffing times. The Court describes the study as follows:

[The] evidence included employee testimony, video recordings of donning and doffing at the plant, and, most important, a study performed by an industrial relations expert, Dr. Kenneth Mericle. Mericle conducted 744 videotaped observations and analyzed how long various donning and doffing activities took. He then averaged the time taken in the observations to produce an estimate of 18 minutes a day for the cut and retrim departments and 21.25 minutes for the kill department.⁴²

Mericle's averages were then used in the Fox study and combined with Tyson's records of gang-time and K-code time to determine which employees worked overtime and how much uncompensated overtime they were due.⁴³ Fox estimated the total amount of uncompensated overtime at \$6.7 million for all employees in the aggregate.⁴⁴ He did this by adding Mericle's average donning and doffing time to Tyson's record of gang-time for each employee, subtracting any K-code time, and then summing up the results over all the employees.⁴⁵ The Supreme Court gave two helpful examples to illustrate the methodology:

For example, if an employee in the kill department had worked 39.125 hours of gang-time in a 6-day workweek and had been paid an hour of K-code time, the estimated number of

39. *See id.*

40. *Id.* at 1043.

41. *See id.* at 1043–44. Other FLSA cases decided before *Tyson Foods* relied on similar studies. *See, e.g.,* *Perez v. Mountaire Farms, Inc.*, 650 F.3d 350, 362 (4th Cir. 2011) (describing expert time studies used to measure the average donning and doffing time in a poultry processing plant).

42. *Tyson Foods*, 136 S. Ct. at 1043.

43. *Id.* at 1043–44.

44. *Id.* at 1044.

45. *Id.* at 1043–44.

compensable hours the employee worked would be: 39.125 (individual number of gang-time hours worked) + 2.125 (the average donning and doffing hours for a 6-day week, based on Mericle's estimated average of 21.25 minutes a day) - 1 (K-code hours) = 40.25 . That would mean the employee was being undercompensated by a quarter of an hour of overtime a week, in violation of the FLSA. On the other hand, if the employee's records showed only 38 hours of gang-time and an hour of K-code time, the calculation would be: $38 + 2.125 - 1 = 39.125$. Having worked less than 40 hours, that employee would not be entitled to overtime pay and would not have proved an FLSA violation.⁴⁶

The key thing to note about these estimates is that they are all based on average donning and doffing times and not on actual times for each employee. If all employees took identical time to don and doff, average time would be the same as actual time. But the evidence made clear that employees took different times, and those differences, though small, significantly affected Tyson's liability to individual employees.⁴⁷ If an employee did not work overtime—and a few minutes per day compiled over a week of work determined that fact—Tyson was not liable to that employee.⁴⁸ These points may seem obvious, but it is important to keep them clearly in mind as we unpack the Court's opinion.

The district court certified a collective action for the FLSA claims and a Rule 23(b)(3) class action for the IWPCCL claims.⁴⁹ The case went to trial. In addition to submitting the Mericle and Fox studies, the plaintiffs also introduced evidence to prove Tyson's plant-wide compensation policies and practices, called a number of employees as witnesses to testify to the nature of the equipment and their own donning and doffing times, and submitted video recordings of donning and doffing activity.⁵⁰ The jury found Tyson liable and

46. *Id.* at 1044.

47. *See id.* at 1055 (Thomas, J., dissenting); *infra* notes 80–81 and accompanying text.

48. *See Tyson Foods*, 136 S. Ct. at 1043 (“Since the employees’ claims relate only to overtime, each employee had to show he or she worked more than 40 hours a week, inclusive of time spent donning and doffing, in order to recover.”).

49. *Id.* A total of 444 employees opted into the FLSA collective action, and a total of 3,344 employees remained in the IWPCCL class at the close of the opt-out period. *Tyson Foods*, 136 S. Ct. at 1043.

50. *See Bouaphakeo v. Tyson Foods, Inc.*, No. 5:07-CV-04009, 2012 WL 4471119, at *3 (N.D. Iowa Sept. 26, 2012) (noting that some employees “testified to the[ir] general practices . . . regarding the donning and doffing of [personal protective equipment (“PPE”)]” including “which PPE items were used by different groups of

returned a verdict in the aggregate amount of roughly \$2.9 million.⁵¹ Tyson moved to set aside the verdict, arguing, in part, that neither the FLSA collective action nor the Rule 23 class action was properly certified.⁵² The district court denied Tyson's motion, and the Eighth Circuit affirmed.⁵³ The Supreme Court granted certiorari and affirmed by a six-to-two margin.

B. *The Court's Holding*

The Supreme Court addressed two issues: (1) whether there was sufficient commonality among the claims to support certification of a class action and a collective action, and (2) whether it was permissible to distribute an aggregate damages award to the class when some class members suffered no legal injury.⁵⁴

As to the first issue, Tyson argued that donning and doffing times varied too much across departments and across employees in the same department to satisfy the FLSA's "similarly situated" requirement and Rule 23(b)(3)'s predominance requirement.⁵⁵ Furthermore, to establish liability, *each* employee had to prove that he or she worked overtime and was not properly compensated for it.⁵⁶ Tyson insisted that these were individual questions that varied from employee to employee—some employees worked overtime and others did not, and those who worked overtime did so for different periods of time.⁵⁷

employees . . . how often people donned and doffed these items, how these items were stored, cleaned, and sanitized, and how long it generally took for people to don and doff the items[.]" and "[w]itnesses also testified that Plaintiffs spent more time donning and doffing their PPE than Defendant paid them in K-code time"), *aff'd*, 765 F.3d 791 (8th Cir. 2014), *aff'd and remanded*, 136 S. Ct. 1036 (2016).

51. *Tyson Foods*, 136 S. Ct. at 1044. Tyson requested a single trial of liability and damages and an aggregate verdict, rather than a bifurcated proceeding and individual damage verdicts for each employee. *Id.* However, Tyson's choice makes no difference to the statistical adjudication issue. Mericle's average figures would still have to be used whether damages were individualized or aggregated.

52. *Id.*

53. *Id.*

54. *See id.* at 1041.

55. *Id.* at 1046. Section 216 of the FLSA requires that all employees in a collective action be "similarly situated[.]" 29 U.S.C. § 216 (2012), and Rule 23(b)(3) requires that common questions "predominate" over individual questions, FED R. CIV. P. 23(b)(3).

56. *See Tyson Foods*, 136 S. Ct. at 1046 ("To be entitled to recovery . . . each employee must prove that the amount of time spent donning and doffing, when added to his or her regular hours, amounted to more than 40 hours in a given week. Petitioner argues that these necessarily person-specific inquiries into individual work time predominate over the common questions raised by respondents' claims, making class certification improper.").

57. *Id.* at 1041.

The Court, however, held that a jury could rely on Mericle's sample averages to determine how many employees worked overtime and how much overtime they worked.⁵⁸ More precisely, the Court held that the sample average supported a rebuttable presumption of each employee's actual time.⁵⁹ Since Tyson did not keep records, it had no evidence to rebut the presumption on an individualized basis, so its "primary defense was to show that Mericle's study was unrepresentative or inaccurate."⁶⁰ For this reason, what had been individual questions about employee work time became common questions about the adequacy of the Mericle study. As a result, both the predominance and similarly situated requirements were satisfied, and the Court affirmed certification of both the class action and the collective action.⁶¹

The second issue arose because the class included employees who never worked overtime and thus had no legal claims. Tyson argued that it was improper to distribute shares of the aggregate damage award to those uninjured employees and that since it was impossible to identify uninjured employees in order to separate them out, the suit could not proceed as a class action.⁶² The Court rejected this argument as premature.⁶³ It held that the district judge should first be given a chance to see if a distribution method could be devised that excludes the uninjured employees.⁶⁴

58. *See id.* at 1047.

59. *See id.* (citing *Anderson v. Mt. Clemens Pottery Co.*, 328 U.S. 680, 685, 687–88 (1946), *superseded by statute*, Act of May 14, 1947, ch. 52, sec. 4, 61 Stat. 84, 86–87 (codified as amended at 29 U.S.C. § 254(a) (2012))).

60. *Id.*

61. *See id.* at 1045–49. In fact, the Court assumed that a FLSA collective action would be permissible if a Rule 23 class action were permissible and therefore focused exclusively on Rule 23 requirements. *See id.* at 1045. There were other common questions in the case, such as whether donning and doffing activities were compensable work under the FLSA and whether the time was excluded under the FLSA's *de minimis* exception, but the Court did not rely on these questions alone to satisfy predominance. *See id.* at 1043, 1045–49.

62. Brief of Petitioner at 51–52, *Tyson Foods, Inc. v. Bouaphakeo*, 136 S. Ct. 1036 (2016) (No. 14-1146). Tyson had originally argued that a class action was proper only if the plaintiffs could prove that the class included only members who actually suffered legal injury. *Tyson Foods*, 136 S. Ct. at 1049 (citing *Petition for Writ of Certiorari at i*, *Tyson Foods, Inc. v. Bouaphakeo*, 135 S. Ct. 2806 (2015) (No. 14-1146)). But it abandoned that argument in favor of a narrower one questioning the distribution of the aggregate award. *See id.* (citing Brief for Petitioner, *supra*, at 49).

63. *Tyson Foods*, 136 S. Ct. at 1050.

64. *Id.* The Court mentioned a possible approach, which, like its liability analysis, assumed that the jury actually found that all employees took an *identical* time to don and doff:

C. *The Court's Reasoning*

One of the most striking aspects of the opinion is how challenging the Court found it to justify the sample average. This is especially surprising because the case seems a particularly strong one for statistical adjudication. Tyson employees were unable to prove their individual overtime with direct evidence because Tyson kept no records of actual donning and doffing times.⁶⁵ Moreover, most, if not all, employees had too little at stake to justify the cost of an individual suit, so a class or collective action was necessary to vindicate the substantive rights at stake.⁶⁶ And the requirements for certifying a class or collective action could be satisfied only with a collective method of determining liability and damages, which statistical sampling supplied.⁶⁷ As for Tyson, its substantive liability would be about the same with sampling as without. As long as the sampling procedure was properly designed—and Tyson raised no objections to its design⁶⁸—aggregate damages based on the sample average would closely approximate total damages based on individual trials.⁶⁹

The following discussion critically reviews the Court's reasoning. In what is perhaps the most important step of its argument, the Court characterizes the case as one involving statistical evidence and a jury inference from sampling results, rather than statistical adjudication

[B]y working backwards from the damages award, and assuming each employee donned and doffed for an identical amount of time (an assumption that follows from the jury's finding that the employees suffered equivalent harm under the policy), it may be possible to calculate the average donning and doffing time the jury necessarily must have found, and then apply this figure to each employee's known gang-time hours to determine which employees worked more than 40 hours.

Id.

65. *Id.* at 1042. And, of course, neither did the employees.

66. Apparently, employees incurred "overtime damages ranging from a few cents to several thousand dollars," and 509 workers suffered injuries ranging from twenty-seven cents to less than \$100. *See Bouaphakeo v. Tyson Foods, Inc.*, 765 F.3d 791, 804 (8th Cir. 2014), *aff'd and remanded*, 136 S. Ct. 1036 (2016). This level of recovery is generally inadequate to support individual litigation. *See Espenscheid v. DirectSat USA, LLC*, 705 F.3d 770, 776 (7th Cir. 2013) (noting in a FLSA case that losses of "several thousand dollars" are not "enough to finance a modern federal lawsuit").

67. *See Tyson Foods*, 136 S. Ct. at 1047.

68. *Id.* at 1044.

69. This result follows from the statistical property that the sample average is close to the population average for a large enough sample. For a discussion of this and other statistical properties of the sample average, see RICHARD J. LARSEN & MORRIS L. MARX, AN INTRODUCTION TO MATHEMATICAL STATISTICS AND ITS APPLICATIONS § 3.9, at 162–73 (2d ed. 1986).

and the application of the sample average directly. Having characterized the case in this way, the Court then justifies the use of the inference to determine liability as well as damages. Finally, the Court removes the *Wal-Mart* barrier to sampling by reinterpreting *Wal-Mart*'s "Trial by Formula" holding.⁷⁰

1. Evidentiary Inference or Statistical Adjudication?

The Court treats the sample averages generated by Mericle's study as so-called "representative evidence" from which a jury could infer the factual findings necessary to decide each employee's separate case.⁷¹ In *Tyson Foods*, each employee's FLSA right was an individual right, and Tyson's liability was owed to each employee individually, not to all employees as a group. In keeping with this individualized focus, the Court framed the "central dispute in the case" as whether the jury could properly infer that "each employee donned and doffed for the same average time observed in Mericle's sample."⁷²

This is an odd way to frame the issue. It was not possible on the facts for a reasonable jury to infer that each employee took the same time to don and doff. Such an inference defies common sense given the obvious differences among employees, and the evidence in the case indicates otherwise.⁷³ Even employees who actually testified at trial described different donning and doffing times.⁷⁴ At one point,

70. See *Tyson Foods*, 136 S. Ct. at 1048.

71. See *id.* at 1046. The phrase "representative evidence" and its close cousin "representative testimony" are standard in FLSA litigation used to refer to the practice of relying on evidence from some employees to support findings for all employees. See, e.g., *Reich v. S. New Eng. Telecomms. Corp.*, 121 F.3d 58, 66–68 (2d Cir. 1997) ("representative testimony" in a FLSA action brought by the secretary of labor); *Indergit v. Rite Aid Corp.*, 52 F. Supp. 3d 522, 525 (S.D.N.Y. 2014) ("representative testimony" in a private FLSA action); *LaFleur v. Dollar Tree Stores, Inc.*, 30 F. Supp. 3d 463, 473–75 (E.D. Va. 2014) (referring to evidence from a sample of employees as "representative testimony" and noting that "[t]estimony of a representative sampling of plaintiffs in a collective action is a procedure often used in FLSA actions"); *Falcon v. Starbucks Corp.*, 580 F. Supp. 2d 528, 540 (S.D. Tex. 2008) (noting that plaintiffs can use "representative testimony" in a FLSA case).

72. *Tyson Foods*, 136 S. Ct. at 1046. The Court assumes throughout the opinion that the jury found identical times. See *id.* at 1050.

73. *Id.* at 1055 (Thomas, J., dissenting) (noting that "cut and retrim employees took between 0.583 minutes and over 10 minutes to don preshift equipment" and "[n]o two employees performed the same activity in the same amount of time").

74. *Id.* at 1057 ("For instance, Mericle's study estimated that kill department employees took an average 6.4 minutes to don equipment at their lockers before their shift—but employee Donald Brown testified that this activity took him around 2 minutes. Others also testified to donning and doffing times that diverged markedly from Mericle's

the Court hedges a bit by referring to “roughly equal time[s].”⁷⁵ But roughly equal is not the same as identical.

This is important because it means that the only thing the jury could have done is use the sample average itself as the time for each and every individual employee.⁷⁶ This strongly suggests that the case is more like one involving statistical adjudication than one involving statistical evidence. If the sample average functioned only as statistical evidence, the jury would have had no reason to rely on it for cases where there was individual evidence, as there was for some of the sampled employees in the Mericle study.⁷⁷ Moreover, as explained further below, if the goal had been to use the available evidence to get as accurate a decision as practically feasible in each individual case, it would have made sense for the Court to require each employee to submit her own average time based on a Mericle-type study of her personal donning and doffing activity.⁷⁸ Although these individualized averages might differ somewhat from the actual historical times, they should be more accurate for individual cases than a population-wide sample average. However, if the Court had required individualized averages, donning and doffing time would have been an individual issue, threatening Rule 23(b)(3) predominance and class action treatment.⁷⁹

I am not suggesting that a jury could never rely on a sample average to infer facts about individual employees. Had Tyson’s liability turned on what was an objectively “reasonable” time to don and doff, then the sample average might have been justified as statistical evidence probative of an objectively “reasonable” time. But the Court treats the sample average as probative of actual time, not

estimates.”); *see also* *Bouaphakeo v. Tyson Foods, Inc.*, No. 5:07-CV-04009-JAJ, 2011 WL 3421541, at *5 (N.D. Iowa Aug. 4, 2011).

75. *See Tyson Foods*, 136 S. Ct. at 1049 (majority opinion).

76. Or the jury could have used a modified version of the sample average. In fact, the jury awarded less than half the total damages that Fox calculated using the sample average. *Id.* at 1044. This means that the jury must have modified the result in some way or discounted some of the claims (or just felt that the actual damages were too high). Unfortunately, there is no way to determine exactly what the jury did. *See id.* at 1052 (Roberts, C.J., concurring) (speculating on what the jury might have done).

77. *See id.* at 1055 (Thomas, J., dissenting) (noting the wide variation in the times taken by the sampled employees observed by Mericle).

78. *See infra* notes 135–36 and accompanying text.

79. In effect, the Court assumes that the jury decided in two steps: that it first inferred actual time for an employee (call her X) from Mericle’s averages and then determined liability and damages for X based on X’s actual time. However, the jury could only have decided in one step: it must have determined liability and calculated damages for X directly on the basis of Mericle’s average without inferring anything about her actual time.

reasonable time. Also, an inference from the sample average to actual times might be acceptable if there were other evidence in the case suggesting that employees took identical times. But Mericle's observations, the video evidence, and the trial testimony all showed that Tyson employees took different times.⁸⁰ To be sure, the differences are quite small—measured in minutes per day—but those minutes make a substantial difference to Tyson's liability as well as to the damages it owed individual employees.⁸¹

In addition, the sample average might have evidentiary value if it supplemented individualized evidence of employee-specific times in some probative way. But this is not how the *Tyson Foods* Court uses it. The Court instead holds that the sample average itself can support a factual determination of individual donning and doffing time.⁸² This holding is far from obvious. Courts and commentators have long struggled with the propriety of using aggregate statistics and base rate probabilities to prove individualized liability elements in the absence of more particularized, case-specific evidence—at least when the substantive law does not clearly invite a statistical approach.⁸³ Indeed,

80. See *Tyson Foods*, 136 S. Ct. at 1055. For a case that refused to allow testimony from a sample of employees to establish overtime compensation and minimum wage violations for other employees in part because employee-specific facts varied too much, see *Espenscheid v. DirectSat USA, LLC*, 705 F.3d 770, 774–75 (7th Cir. 2013).

81. Indeed, slight deviations in the sample average had a substantial impact on the number of employees with valid overtime claims, indicating that small differences in donning and doffing times were significant. See *Tyson Foods*, 136 S. Ct. at 1055 (“If Mericle’s averages even slightly overestimated average donning and doffing times, another 282 class members would have no overtime claims. If average donning or doffing times dropped from 18–21 minutes to 15 minutes, Fox stated, another 110 employees had no overtime claims.”).

82. See *id.* at 1046–47 (majority opinion) (noting that because employees could use the sample average to prove their times in individual suits, “that sample is a permissible means of establishing the employees’ hours worked in a class action”). Indeed, if each plaintiff had to supplement the sample average with individual evidence of donning and doffing activity, it would have been very difficult for named plaintiffs to satisfy Rule 23(b)(3)’s predominance requirement. The sufficiency of the sample average ensured that individual donning and doffing times could be proved on a class-wide basis and thus facilitated a predominance finding. See *supra* text accompanying notes 55–61.

83. See *Howard v. Wal-Mart Stores, Inc.*, 160 F.3d 358, 359–60 (7th Cir. 1998) (discussing some of the problems with relying exclusively on a base rate statistic to prove an individual element, at least when it would be reasonable to expect case-specific evidence); *Smith v. Rapid Transit, Inc.*, 58 N.E.2d 754, 755 (Mass. 1945) (rejecting the sufficiency of the statistical evidence offered to identify the owner of a bus that caused an accident); Glen O. Robinson & Kenneth S. Abraham, Essay, *Collective Justice in Tort Law*, 78 VA. L. REV. 1481, 1486–87 (1992) (criticizing tort law’s individual, case-specific focus and noting in particular “the traditional hostility of courts to statistical and probabilistic evidence, which is aggregative in character, in favor of individual, ‘clinical’-

the *Tyson Foods* Court itself recognized that the evidentiary sufficiency of the sample average requires special justification. For that purpose, the Court relied on the fact that Tyson failed to keep records of individual times as required by the FLSA, that its failure deprived employees of the ability to prove their actual times precisely, and that allowing use of the sample average filled the evidentiary gap and facilitated enforcement of the FLSA's important substantive policies.⁸⁴

Rather than trying a sample and averaging the sample verdicts, the trial judge in *Tyson Foods* relied on experts to find the facts in a sample of cases and allowed the jury to modify the sample average in light of testimonial and video evidence.⁸⁵ But the key point is the same: the sample average is what controlled the outcome, not individualized fact finding, and as a result, some employees received more than their substantive entitlements and some received less.⁸⁶

It is true that the Court uses the sample average to support a rebuttable rather than a conclusive presumption of actual donning and doffing times.⁸⁷ However, since Tyson kept no records of donning and doffing time, it could not rebut the presumption. So for all practical purposes, the presumption was conclusive.

type evidence"); see also Edward K. Cheng, *Reconceptualizing the Burden of Proof*, 122 YALE L.J. 1254, 1269 & n.30, 1271 (2013) (noting that "courts have generally rejected awarding damages to the plaintiff based on . . . 'naked' statistical evidence" and tend to favor "individualized 'direct' proof"); Chris Guthrie, Jeffrey J. Rachlinski & Andrew J. Wistrich, *Inside the Judicial Mind*, 86 CORNELL L. REV. 777, 810 (2001) (noting the preference of judges for individuating evidence over statistical evidence); Laurence H. Tribe, *Trial by Mathematics: Precision and Ritual in the Legal Process*, 84 HARV. L. REV. 1329, 1343-44, 1377 (1971) (noting that courts have been reluctant to rely on statistical evidence as sufficient alone and recommending great caution in the use of mathematical proof); Richard W. Wright, *Causation, Responsibility, Risk, Probability, Naked Statistics, and Proof: Pruning the Bramble Bush by Clarifying the Concepts*, 73 IOWA L. REV. 1001, 1050-51 (1988) (noting that "[j]udges generally have refused to accept naked statistics or *ex ante* causal probabilities as evidence of what actually happened on a particular occasion" and observing that "[t]hey instead have insisted on particularistic evidence and *ex post* causal probabilities based on such evidence").

84. See *Tyson Foods*, 136 S. Ct. at 1047; see also *infra* notes 121-127 and accompanying text.

85. See *Tyson Foods*, 136 S. Ct. at 1043-44. This is similar to the way the court structured statistical adjudication in *Hilao*. See *Hilao v. Estate of Marcos*, 103 F.3d 767, 782-84 (9th Cir. 1996); see also *supra* notes 21-24 and accompanying text.

86. *Tyson Foods* is unusual in going to trial; most class actions settle. The fact that it went to trial and that Tyson's motion was filed after the trial verdict allowed the Court to defer heavily to the jury's decision, or more precisely, to its characterization of the jury's decision. It is worth noting, however, that in most cases, the permissibility of statistical adjudication will have to be decided much earlier, at the initial stage of class certification.

87. See *Tyson Foods*, 136 S. Ct. at 1047.

To be sure, when population variance is small, the sample average can be a reasonably good approximation of actual times. But this evidentiary benefit is not enough alone to justify shifting the burden to the defendant where there is case-specific evidence to support more accurate individual results or where individualized sampling is possible and likely to produce better estimates. In other words, there must be some special reason that justifies the use of aggregate sampling, regardless of whether it creates a conclusive or a rebuttable presumption. In *Tyson Foods*, that reason had to do with the evidentiary obstacle to enforcement created by the defendant's failure to keep adequate records and the importance of enabling private enforcement of the FLSA's substantive policies.⁸⁸ Thus, *Tyson Foods* is similar to other statistical adjudication cases—the sample average is justified for its enforcement benefits and not merely for its probative value.

2. Inference of Liability and Damages?

Having characterized the issue in terms of evidentiary inference, the Court then turned to the task of justifying the inference. To do this, it relied on a 1946 decision, *Anderson v. Mt. Clemens Pottery Co.*⁸⁹ The Court read *Mt. Clemens* to hold that a rebuttable inference of liability and damages based on the sample average was permissible in individual FLSA suits whenever the employer failed to keep adequate records of employee time.⁹⁰ From this, it concluded that the

88. See *infra* Sections III.A.1.–2.

89. 328 U.S. 680 (1946), *superseded by statute*, Act of May 14, 1947, ch. 52, sec. 4, 61 Stat. 84, 86–87 (codified as amended at 29 U.S.C. § 254(a) (2012)). *Mt. Clemens* was a representative suit for violation of the FLSA brought by a union and seven of its members on behalf of all similarly situated employees, in which about 300 employees and ex-employees authorized the plaintiffs to represent them. *Mt. Clemens Pottery Co. v. Anderson*, 149 F.2d 461, 461 (6th Cir. 1945), *rev'd*, 328 U.S. 680 (1946).

90. *Tyson Foods*, 136 S. Ct. at 1047. Actually, it is not exactly clear how the district judge calculated the average time estimates in *Mt. Clemens*. The judge referred the case to a special master who, after hearing testimony from a number of witnesses, concluded that the computation of overtime would be too speculative to support recovery. See *Mt. Clemens*, 149 F.2d at 463–64. The district judge disagreed with the special master's conclusion and created a formula based on what the Sixth Circuit refers to as an “estimated average of overtime worked.” See *id.* at 464–65. But it is not clear how the district judge calculated this “estimated average.” In the end, the Sixth Circuit held that “the arbitrary formula applied by the district judge, in lieu of acceptance of the master's findings, produced a judgment based upon surmise and conjecture, which cannot be sustained.” *Id.* at 465. But the Supreme Court reversed. *Mt. Clemens*, 328 U.S. at 687. It held that when the employer fails to keep adequate time records, employees can prove their individual times by evidence sufficient to support a “just and reasonable inference.”

same inference applied to a class action aggregating those individual FLSA claims.⁹¹

The trouble with relying on *Mt. Clemens* is that it did not go as far as the *Tyson Foods* Court assumes. The *Mt. Clemens* Court authorized the use of sampling to estimate damages, but only after liability was already established:

In such a situation [where the employer fails to keep statutorily mandated records,] we hold that an employee has carried out his burden *if he proves that he has in fact performed work for which he was improperly compensated* and if he produces sufficient evidence to show the amount and extent of that work as a matter of just and reasonable inference. . . . Nor is such a result to be condemned by the rule that precludes the recovery of uncertain and speculative damages. That rule applies only to situations where the fact of damage is itself uncertain. *But here we are assuming that the employee has proved that he has performed work and has not been paid in accordance with the statute.* The damage is therefore certain. The uncertainty lies only in the amount of damages arising from the statutory violation by the employer.⁹²

It is more difficult to justify the use of sampling to determine liability.⁹³ I shall argue below that stretching *Mt. Clemens* to cover

Id. But it also remanded “for the determination of the amount of walking time involved and the amount of preliminary activities performed, giving due consideration to the *de minimis* doctrine and calculating the resulting damages under the Act.” *Id.* at 694.

91. See *Tyson Foods*, 136 S. Ct. at 1046–47. The *Mt. Clemens* rule allowing sampling is, of course, a rule of federal law, raising the question of why it applies to the state-created IWPC claim. The most likely answer is that the IWPC claim is a hybrid state-federal claim that incorporates the rule. The IWPC statute requires an employer to “pay all wages due its employees” but does not specify what wages are due. IOWA CODE ANN. § 91A.3 (West, Westlaw through 2016 Legis. Sess.). The plaintiffs in *Tyson Foods* relied on the FLSA to fill the gap. As the district judge noted, this made the IWPC claim “essentially ‘duplicative’ of the FLSA claim.” *Bouaphakeo v. Tyson Foods, Inc.*, 564 F. Supp. 2d 870, 883–84 (N.D. Iowa 2008). It follows that the *Mt. Clemens* rule, as a substantive component of the FLSA right, applies to the IWPC claim as well:

The parties do not dispute that, in order to prove a violation of the Iowa statute, the employees had to do no more than demonstrate a violation of the FLSA. In this opinion, then, no distinction is made between the requirements for the class action raising the state-law claims and the collective action raising the federal claims.

Tyson Foods, 136 S. Ct. at 1045.

92. *Mt. Clemens*, 328 U.S. at 687–88 (emphasis added).

93. It is worth noting, however, that some lower courts, before *Tyson Foods*, had extended *Mt. Clemens* to proof of liability as well as damages in FLSA cases. See, e.g.,

liability makes sense as a policy matter. But it is significant to note here that the *Tyson Foods* majority fudged the issue rather than openly concede what it was doing and explicitly justify the broader use on legal and policy grounds.⁹⁴

3. Trial by Formula?

The *Tyson Foods* Court also had to deal with *Wal-Mart's* rejection of sampling as “Trial by Formula.”⁹⁵ This was not easy to do because the use of statistical adjudication in *Tyson Foods* was structurally very similar to its use in *Wal-Mart*. *Wal-Mart* was a putative class action brought on behalf of roughly 1.5 million female employees of Wal-Mart Stores nationwide.⁹⁶ The class alleged Title VII disparate treatment claims based on a company-wide policy of gender discrimination in pay and promotion.⁹⁷ The district court certified a Rule 23(b)(2) class action for injunctive, declaratory, and backpay relief, and the Ninth Circuit affirmed.⁹⁸ With respect to backpay, the Ninth Circuit suggested that the district court might use a sampling procedure that involved appointing a special master to

Garcia v. Tyson Foods, Inc., 770 F.3d 1300, 1307 (10th Cir. 2014) (allowing representative evidence to prove liability in a FLSA case). Still, it is much easier to justify the use of sampling at the damages stage. It is well established that a plaintiff need only propose a method for estimating damages that permits a “‘just and reasonable inference’ . . . even though the result is only an approximation.” *Cristallina S.A. v. Christie, Manson & Woods Int’l, Inc.*, 502 N.Y.S.2d 165, 173 (App. Div. 1986) (quoting *Story Parchment Co. v. Paterson Parchment Paper Co.*, 282 U.S. 555, 563 (1931)). One reason for this liberal approach is the unfairness of imposing the risk of uncertain measurement on a plaintiff when the defendant is a proven wrongdoer.

94. This is especially striking in light of Justice Thomas’s dissent, which takes the majority to task for misinterpreting *Mt. Clemens* in precisely this way. *Tyson Foods*, 136 S. Ct. at 1058 (Thomas, J., dissenting).

95. See *Wal-Mart Stores, Inc. v. Dukes*, 564 U.S. 338, 367 (2011).

96. *Id.* at 342.

97. *Id.* at 343.

98. See *id.* at 346–47. Rule 23(b)(2) authorizes a class action for injunctive and declaratory relief when the defendant “has acted or refused to act on grounds that apply generally to the class, so that final injunctive relief or corresponding declaratory relief is appropriate respecting the class as a whole.” FED. R. CIV. P. 23(b)(2). While the Rule says nothing about monetary relief, lower courts before *Wal-Mart* routinely allowed backpay claims under the Rule 23(b)(2) umbrella on the theory that backpay was merely incidental to Rule 23(b)(2) injunctive relief. See, e.g., *Thorn v. Jefferson-Pilot Life Ins. Co.*, 445 F.3d 311, 332 (4th Cir. 2006) (“[O]ur prior cases have held that Rule 23(b)(2) class certification is proper in the Title VII context . . . because injunctive or declaratory relief predominates despite the presence of a request for back pay.”); *Pettway v. Am. Cast Iron Pipe Co.*, 494 F.2d 211, 257 (5th Cir. 1974) (“Back pay has been granted by courts in (b)(2) class actions where injunctive or declaratory relief has also been requested . . .”).

depose a random sample of class members and extrapolating from the sample results to an aggregate backpay award.⁹⁹

The Supreme Court reversed on the ground that the named plaintiffs had failed to demonstrate sufficient commonality for class certification.¹⁰⁰ The Court also held that a Rule 23(b)(2) class action could not include backpay claims and criticized the Ninth Circuit's proposed sampling procedure.¹⁰¹ With respect to the Rule 23(b)(2) backpay issue, the Court explained that in a Title VII case like *Wal-Mart*, each class member has a presumptive entitlement to backpay once the class succeeds in proving a pattern and practice of discrimination.¹⁰² But Title VII then gives the employer the right to rebut the presumption for each employee if it can show lawful reasons for the employee's particular treatment.¹⁰³ The Supreme Court concluded that the Ninth Circuit's sampling procedure deprived the defendant of its right to rebut the presumption in each individual case—i.e., its right “to litigate its statutory defenses to individual claims”—and thus altered the substantive law in violation of the REA.¹⁰⁴

The similarities to *Tyson Foods* are striking. In both cases, most class members had claims for monetary relief that were too small to support individual suits and therefore the only realistic way to provide relief was through a class action or other form of aggregation. In both cases, aggregate treatment was possible only if the court could calculate an aggregate award on a class-wide basis. In both cases, courts either proposed or implemented a form of statistical

99. *Wal-Mart Stores*, 564 U.S. at 367. The Ninth Circuit described its proposed approach as follows:

A sample set of the class members would be selected, as to whom liability for sex discrimination and the backpay owing as a result would be determined in depositions supervised by a master. The percentage of claims determined to be valid would then be applied to the entire remaining class, and the number of (presumptively) valid claims thus derived would be multiplied by the average backpay award in the sample set to arrive at the entire class recovery—without further individualized proceedings.

Id. (citing *Dukes v. Wal-Mart Stores, Inc.*, 603 F.3d 571, 625–27 (9th Cir. 2010), *rev'd*, 564 U.S. 338 (2011)). This proposal followed the approach used in *Hilao*. See *Hilao v. Estate of Marcos*, 103 F.3d 767, 782–84 (9th Cir. 1996); see also *supra* note 21–24 and accompanying text.

100. *Wal-Mart Stores*, 564 U.S. at 359.

101. See *id.* at 366–67.

102. See *id.*

103. *Id.*

104. *Id.*

adjudication to do this. And in both cases, the sample average was used to determine liability as well as damages.¹⁰⁵

Despite these similarities, the *Tyson Foods* Court managed to distinguish *Wal-Mart*. To do so, it focused on the heterogeneity of the two employee classes. According to the Court, the wide differences in employee circumstances across all Wal-Mart stores made the nationwide class too heterogeneous to permit the use of a sample average.¹⁰⁶ In *Tyson Foods*, by contrast, “each employee worked in the same facility, did similar work, and was paid under the same policy.”¹⁰⁷ This made the Tyson employee class much less heterogeneous, which justified application of the *Mt. Clemens* rule and allowed reliance on the sample average.¹⁰⁸

There are two notable problems with this line of reasoning. First, it does not match what the Court actually said in *Wal-Mart*. In *Wal-Mart*, the problem with sampling was that it deprived the company of its right “to litigate its statutory defenses to individual claims.”¹⁰⁹ Given this, it is a bit surprising that the *Tyson Foods* Court relied on intra-class heterogeneity to distinguish *Wal-Mart* when it could have relied on the fact that a rebuttable presumption allows the defendant to present its defenses. Perhaps the Court worried that its supposedly rebuttable presumption was not actually rebuttable after all and that using the sample average to shift the burden on an element of the prima facie case when the defendant has no way to rebut the presumption is not all that different from cutting off a defense.¹¹⁰ In

105. See *Tyson Foods, Inc. v. Bouaphakeo*, 136 S. Ct. 1036, 1046–47 (2016); *Wal-Mart Stores*, 564 U.S. at 367. In *Wal-Mart*, the liability issue was whether Wal-Mart had lawful reasons for treating a particular employee the way it did. See *Wal-Mart Stores*, 564 U.S. at 366–67. In *Tyson Foods*, the liability issue was whether any given employee actually worked overtime. See *Tyson Foods*, 136 S. Ct. at 1047.

106. See *Tyson Foods*, 136 S. Ct. at 1048. The Court reasoned that this intra-class heterogeneity precluded the use of the sample average in an individual case, and then concluded that the Rules Enabling Act (“REA”) precluded its use in a class action as well. *Id.* (“Permitting the use of that sample in a class action, therefore, would have violated the Rules Enabling Act by giving plaintiffs and defendants different rights in a class proceeding than they could have asserted in an individual action.”).

107. *Id.*

108. Actually, the Court argues that since *Mt. Clemens* allowed the sample average to be used in an individual suit, it must also be available in a class action, or else there would be a REA violation. See *id.* at 1047–48.

109. *Wal-Mart Stores*, 564 U.S. at 367.

110. The *Tyson Foods* Court noted that “[s]ince there were no alternative means for the employees to establish their hours worked [besides the sample average], [Tyson’s] primary defense was to show that Mericle’s study was unrepresentative or inaccurate.” *Tyson Foods*, 136 S. Ct. at 1047. In other words, there was no point in giving Tyson a

any event, it is not clear what heterogeneity has to do with a right to litigate individual defenses. As long as the cases exhibit legally relevant differences, *Wal-Mart* would seem to require individual litigation. And legally relevant differences can exist even when a class has low heterogeneity.

Second, *Tyson Foods* confuses *Wal-Mart*'s discussion of Rule 23(a)(2) commonality with its discussion of Rule 23(b)(2) backpay relief.¹¹¹ It is true that the *Wal-Mart* Court emphasized the sprawling and internally heterogeneous nature of the class, but it did so in the course of holding that the plaintiffs failed to provide enough evidence of a company-wide discriminatory policy to satisfy the Rule 23(a)(2) commonality requirement.¹¹² The Rule 23(b)(2) backpay issue is different. It arises only *after* the plaintiffs prove the existence of a company-wide policy of discrimination, since it is only then that the presumptive right to backpay attaches.¹¹³ This makes a huge difference. If *Wal-Mart* actually has a company-wide discriminatory policy affecting pay and promotion, then that policy lends cohesion to the class and makes it more homogeneous than the *Tyson Foods* Court assumes. Indeed, it is not clear which is more heterogeneous: a *Wal-Mart* class subject to a company-wide discriminatory policy or a *Tyson Foods* class with substantial overtime variations.¹¹⁴

Of course, the *Tyson Foods* Court had to do something to disarm *Wal-Mart*, and distinguishing precedent by reinterpreting it is a tried-and-true strategy. But the way the Court distinguished *Wal-Mart* creates a problem for the use of sampling in other aggregations. The Court reasons that sampling is proper in class actions *only* because it is proper in individual suits.¹¹⁵ Thus, the permissibility of sampling in

chance to defend against individual claims because it had nothing to present by way of defense. This brings the *Tyson Foods* case rather close to *Wal-Mart*. The point of *Wal-Mart* is that sample averaging deprives parties of their right to litigate individual issues by averaging over individual facts. *Wal-Mart Stores*, 564 U.S. at 367.

111. For example, the *Tyson Foods* Court says: "The plaintiffs in *Wal-Mart* proposed to use representative evidence as a means of overcoming [the] absence of a common policy [for purposes of determining an aggregate backpay award]." *Tyson Foods*, 136 S. Ct. at 1048. This is simply wrong. The "representative evidence" in *Wal-Mart* was to be used to determine backpay only after plaintiffs proved the existence of a common policy. *See Wal-Mart Stores*, 564 U.S. at 367.

112. *See Wal-Mart Stores*, 564 U.S. at 356–60.

113. *Id.* at 336–37.

114. The existence of a company-wide discriminatory policy is a good reason to believe that most employees suffered from discrimination, whereas the fact that *Tyson* employees worked in the same plant and were paid under the same gang-time system says nothing about their individual donning and doffing times.

115. *See Tyson Foods*, 136 S. Ct. at 1048.

individual suits is a *necessary* condition for the use of sampling in class actions. This limitation ignores important differences between a class action and an individual suit, differences that sometimes make sampling desirable in the former when it might not be desirable in the latter. This point is discussed at greater length in Part III below.

III. THE FUTURE OF SAMPLING AFTER *TYSON FOODS*

The *Tyson Foods* Court expressly disavows “broad and categorical rules” and endorses a case-specific approach.¹¹⁶ It opens the door to broader use of sampling but provides little guidance to lower court judges, who must decide when and how to use it. The following discussion develops factors and principles to guide those decisions. Section A unpacks the factors that the Court actually identifies; Section B addresses sampling’s outcome effects; and Section C discusses some other relevant factors that should be taken into consideration.

A word of clarification is in order at the outset. The following account is not the only reasonable interpretation of the *Tyson Foods* opinion. I offer it as the best principled interpretation. My claim is that the general principles that best fit and justify what the Court says and does in *Tyson Foods* have normative extension and justify sampling in a wider range of cases.

A. *The Court’s Factors*

The *Tyson Foods* Court mentions two factors that it considers relevant to the use of sampling: (1) “the purpose for which the sample is being introduced,” and (2) “the underlying cause of action.”¹¹⁷ In addition, the Court makes clear in its discussion of *Wal-Mart* that the degree of heterogeneity matters, and it also notes that any sampling methodology must be reliable.¹¹⁸

116. *Id.* at 1049.

117. *Id.* (“The fairness and utility of statistical methods in contexts other than those presented here will depend on facts and circumstances particular to those cases.”); *see also id.* at 1046 (“It follows that the Court would reach too far were it to establish general rules governing the use of statistical evidence, or so-called representative evidence, in all class-action cases.”).

118. *Id.* at 1046 (citing FED. R. CIV. P. 401, 403, 702) (stating that the permissibility of a statistical sample “turns . . . on the degree to which the evidence is reliable in proving or disproving the elements of the relevant cause of action”).

1. The Purpose of Sampling

The Court says little about the first factor, and its meaning is far from clear. As we have seen, sampling is sometimes used to generate statistical evidence helpful for determining case-specific facts.¹¹⁹ But this is not how it is used in *Tyson Foods*. The sample average is not just evidence of actual donning and doffing times; it is the time that determines liability for each employee.¹²⁰

When justifying the sample average, the Court relies on Tyson's failure to keep adequate records of donning and doffing times.¹²¹ Given this, one might reasonably read the *Tyson Foods* holding as limited to cases where the sample average is needed to overcome an evidentiary obstacle created by the defendant.¹²² However, I believe

119. See *supra* notes 8–11 and accompanying text.

120. See *Tyson Foods*, 136 S. Ct. at 1046–47. More precisely, the sample average in *Tyson Foods* shifts the burden to the defendant. See *id.* at 1047. In this respect, it might seem analogous to other doctrines that shift the burden on liability issues when rebuttal evidence is unlikely, such as the burden-shifting rule of *Summers v. Tice*, 199 P.2d 1, 4–5 (Cal. 1948), or market share theories in products liability, see *Sindell v. Abbott Labs.*, 607 P.2d 924, 937 (Cal. 1980). These analogies are useful, but they are also different from *Tyson Foods* in important ways. For one thing, the FLSA duty is statutory, and the statute prescribes that liability attaches only if the employee worked overtime. See 29 U.S.C. § 216(b) (2012). The *Summers* burden rule and market share theories are common law doctrines, and judges have more latitude to alter the liability elements of common law torts. Moreover, both of the common law burden-shifting rules apply only when the plaintiff proves that the defendant violated a legal standard owed to her. See *Sindell*, 607 P.2d at 936; *Summers*, 199 P.2d at 3–4. In *Summers v. Tice*, for example, the plaintiff proved that both shooters were negligent; the only question was whose bullet actually hit the plaintiff. See *Summers*, 199 P.2d at 2. Market share theories impose liability based on a probabilistic estimate that the plaintiff used the particular brand, but they presuppose that the product itself is defective or the defendants are otherwise legal wrongdoers. See *Sindell*, 607 P.2d at 925–26, 936–37. To be sure, the Mericle and Fox studies make clear that Tyson failed to pay proper overtime compensation to some employees and thus violated the FLSA as to them. See *Tyson Foods*, 136 S. Ct. at 1043–44. But Tyson did not fail to pay overtime for every employee who benefited from the presumption. In other words, the *Tyson Foods* Court shifts the burden for *all* employees in the class without any proof that the FLSA overtime compensation duty was violated for each and every one. See *id.* at 1047–49.

121. See *Tyson Foods*, 136 S. Ct. at 1047.

122. In a draft article finalized too late for me to respond to here, Jonah Gelbach argues that *Tyson Foods* treats the sample average similar to how courts treat counterfactual evidence in other cases. Jonah B. Gelbach, *The Triangle of Law and the Role of Evidence in Class Litigation*, 165 U. PENN. L. REV. (forthcoming 2017) (manuscript at 9–12, 23), <https://ssrn.com/abstract=2905911> [<https://perma.cc/QVN6-6X53> (staff-uploaded archive)]. I disagree. In the antitrust cases involving counterfactual evidence, for example, statistical evidence is required by the inherent nature of the substantive issue. See *supra* text accompanying note 11. The problem in *Tyson Foods* is different. Tyson employees actually worked the time they did; there is no need to posit a hypothetical state of affairs in order to make sense of the required liability element. The

that a broader interpretation better fits the Court's reasoning. According to this interpretation, the existence of an evidentiary gap is not the crux of the problem; it is just the particular way that the problem manifested itself in the *Tyson Foods* case. Quoting *Mt. Clemens*, the *Tyson Foods* Court emphasized the "remedial nature" of the FLSA, the "great public policy . . . it embodies," and the injustice of leaving an employer free to retain the benefits of employee work without paying for those benefits.¹²³ It never said that these policy concerns are triggered *only* when enforcement is difficult *because of an evidentiary gap*. The existence of an evidentiary gap happened to be the enforcement obstacle in *Tyson Foods* and *Mt. Clemens*, but the reason why the gap mattered—that it frustrated the enforcement of important substantive policies—potentially extends to any obstacle that has the same effect.

In this regard, it is important to bear in mind that the Court never treated Tyson's failure to keep adequate records as a spoliation tort, nor did it limit its holding to situations where the defendant purposefully alters evidence in order to avoid liability.¹²⁴ If Tyson had been guilty of spoliation, shifting the burden might have been justified on moral and evidentiary grounds.¹²⁵ But the Court did not rely on the spoliation doctrine. Indeed, Tyson might have chosen not to keep adequate records simply because it believed—wrongly as it turned out—that its K-code times covered its FLSA obligations and thus that it did not have to incur the cost of keeping track of the daily donning and doffing times for each of its more than three thousand employees.¹²⁶

evidentiary gap does not inhere in the nature of the overtime issue; it arises from Tyson's failure to keep records. Thus, use of the sample average must be justified in a different way.

123. *Tyson Foods*, 136 S. Ct. at 1047 (quoting *Anderson v. Mt. Clemens Pottery Co.*, 328 U.S. 680, 687 (1946), *superseded by statute*, Act of May 14, 1947, ch. 52, sec. 4, 61 Stat. 84, 86–87 (codified as amended at 29 U.S.C. § 254(a) (2012))); *see Anderson*, 328 U.S. at 688 ("[T]he employer, having received the benefits of such work, cannot object to the payment for the work on the most accurate basis possible under the circumstances.").

124. The spoliation doctrine allows a court to presume facts when a party manipulates the evidence so that the opposing party cannot prove the fact. *See Lawrence Solum & Stephen Marzen, Truth and Uncertainty: Legal Control of the Destruction of Evidence*, 36 EMORY L.J. 1085, 1087–94 (1987) (explaining the spoliation inference).

125. *See id.* at 1138–40.

126. *Cf. Tyson Foods*, 136 S. Ct. at 1059 (Thomas, J., dissenting) ("The majority thus puts employers to an untenable choice. They must either track any time that might be the subject of an innovative lawsuit, or they must defend class actions against representative evidence that unfairly homogenizes an individual issue.").

Moreover, it is not even clear that violation of a statutory record-keeping duty is a necessary condition for application of the *Mt. Clemens* rule. Some courts have applied *Mt. Clemens* without a record-keeping violation.¹²⁷ This is consistent with the notion that sampling's purpose is to enforce FLSA rights and secure the statute's compensation and deterrence goals rather than to punish the defendant or incentivize record keeping.

Thus, *Tyson Foods*, broadly interpreted, stands for the proposition that sampling can be used to overcome any serious proof obstacle that systematically deprives a large number of injured parties of compensation, impedes enforcement of the substantive law, and leaves the defendant free to retain the benefits of its unlawful conduct—provided, of course, that sampling is otherwise consistent with the applicable substantive law.¹²⁸ This proposition and the principles it embodies apply not just to FLSA claims but also to the collective litigation of consumer protection, securities fraud, antitrust, and other cases involving small claims.¹²⁹ In cases like these, the

127. See *Villalpando v. Exel Direct Inc.*, Nos. 12-cv-04137-JCS, 13-3091-JCS, 2016 WL 1598663, at *9 (N.D. Cal. Apr. 21, 2016) (construing *Tyson Foods* and *Mt. Clemens* to extend to situations where the employer's failure to keep adequate records does not violate an explicit record-keeping duty). Nor does it matter that the defendant had reason to rely on third parties to keep records. *Id.* (rejecting the defendant's argument that independent contractors should have kept the records). Moreover, the *Mt. Clemens* Court itself noted that its rule applied "even where the lack of accurate records grows out of a bona fide mistake as to whether certain activities or non-activities constitute work." *Mt. Clemens*, 328 U.S. at 688. And a similar principle has been applied in other contexts that do not involve statutory record-keeping violations. See *Melgar v. CSK Auto, Inc.*, No. 13-cv-03769-EMC, 2015 WL 9303977, at *8–9 (N.D. Cal. Dec. 22, 2015) (allowing plaintiffs to rely on self-identification to satisfy the ascertainability requirement for class certification partly because the defendant failed to keep records and "even if there is no statute that explicitly requires recordkeeping for business expenses").

128. It is also important to bear in mind that while overcoming serious proof obstacles supports sampling, there might be contrary principles weighing against its use in particular cases or rules that dictate a different result. For example, applied to most state claims, sampling would almost certainly be considered "substantive" and thus available only if the applicable state law allows it. See *Erie R.R. Co. v. Tompkins*, 304 U.S. 64, 78 (1938); see also *Cimino v. Raymark Indus., Inc.*, 151 F.3d 297, 321 (5th Cir. 1998) (holding *inter alia* that the district court's use of sampling impermissibly altered state substantive law in violation of the *Erie* command).

129. It is worth mentioning that the Supreme Court has been reluctant to embrace this proposition in the arbitration context when negative expected value claims are involved. In *American Express Co. v. Italian Colors Restaurant*, 133 S. Ct. 2304 (2013), for example, the Court held that a class waiver was enforceable in arbitration even though the plaintiffs' federal antitrust claims were too small to justify proceeding individually. *Id.* at 2309–11 (holding that the effective vindication doctrine applies to the "prospective waiver of a party's right to pursue statutory remedies" . . . [b]ut the fact that it is not worth the expense involved in *proving* a statutory remedy does not constitute the elimination of the

defendant's unlawful conduct systematically produces small harms to large numbers of geographically dispersed individuals.¹³⁰ The obstacle to enforcement is not an evidentiary gap, but a negative expected value ("NEV") problem.¹³¹ Most plaintiffs have too little at stake to cover the costs of an individual suit. Aggregation makes suit feasible by attracting an attorney willing to represent the group in return for a percentage of the total recovery. The problem with aggregation, however, is the presence of individual reliance, causation, and damages issues.¹³² Sampling solves this problem by enabling collective resolution of individual issues.

Based on this interpretation, the Court's purpose factor should support the use of statistical adjudication in a case like *Wal-Mart*. The backpay claims in *Wal-Mart* were for small amounts, too small to justify individual suits. Thus, the Ninth Circuit's proposed sampling procedure removed a NEV obstacle that deprived claimants of their entitlement to recover backpay and impeded the enforcement of Title VII rights.¹³³ The Supreme Court invoked the REA to reject the Ninth Circuit's approach,¹³⁴ and I shall discuss that argument later. For now, it is important to note that the purpose factor weighs in favor of sampling in a case like *Wal-Mart*.

right to pursue that remedy" (quoting *Mitsubishi Motors Corp. v. Soler Chrysler-Plymouth, Inc.*, 473 U.S. 614, 637 n.19 (1985))). It is difficult to see how this holding can be justified as a matter of procedural principle detached from the Federal Arbitration Act and the Supreme Court's arbitration jurisprudence, and it is unclear whether and how far the Court will apply it beyond the arbitration context. I am grateful to Maria Glover for alerting me to this point.

130. For a discussion noting the benefits and costs of small claims class actions, see generally Jonathan R. Macey & Geoffrey P. Miller, *The Plaintiffs' Attorney's Role in Class Action and Derivative Litigation: Economic Analysis and Recommendations for Reform*, 58 U. CHI. L. REV. 1 (1991).

131. In other words, the expected cost of litigating an individual suit exceeds the expected recovery, so the suit has a negative expected value. See ROBERT G. BONE, *CIVIL PROCEDURE: THE ECONOMICS OF CIVIL PROCEDURE* 20–36 (2003) (explaining expected value and negative expected value suits).

132. For example, these individual issues can make it difficult to satisfy the predominance requirement for certification of a Rule 23(b)(3) class action. See Robert H. Klonoff, *The Decline of Class Actions*, 90 WASH. U. L. REV. 729, 792–807 (2013). These individual issues might be left to individual suits, but the cost of those suits, even narrowed to just these issues, is likely to be prohibitive given the small amounts at stake.

133. Given the centrality of *Wal-Mart*'s individual defenses to backpay claims, see *Wal-Mart Stores, Inc. v. Dukes*, 564 U.S. 338, 367 (2011), it would have been very difficult to satisfy the predominance requirement for a Rule 23(b)(3) class action without the use of sampling.

134. *Id.*

Tyson Foods is very similar to *Wal-Mart* in this regard. It is actually better understood as a small claims case with a NEV obstacle than a case with an evidentiary gap.¹³⁵ The assumption that plaintiffs cannot prove liability without using the sample average is highly questionable. All Dr. Mericle did was average *current* donning and doffing times for a random sample of *current* employees. He had no access to historical information; Tyson kept no historical records of donning and doffing time. But this means that there was nothing stopping an individual employee from hiring Mericle or some other expert to videotape her own donning and doffing activity and average her personal times. Employee-specific averages would approximate individual donning and doffing times better than the employee-wide sample average used in the case. Thus, it is simply not true that Tyson employees had no alternative but to use Mericle's sample averages. They could have used their own individual averages instead, and those would have produced more accurate liability determinations.¹³⁶

The problem is that many, if not all, Tyson employees had too little at stake to justify the expense of an individual time study. Moreover, employee-specific time studies would make the overtime issue an individual one, which could easily scuttle the aggregate treatment needed to overcome the NEV obstacle. If this is the problem, though, then statistical adjudication serves the same purpose in *Tyson Foods* as in *Wal-Mart* or any other case involving small claims: it removes a NEV obstacle to private enforcement of the substantive law.¹³⁷

135. If Tyson had kept adequate records of donning and doffing times, employees presumably would have had to prove their overtime on an individualized basis. But that would have involved an easy mathematical calculation suitable for a computer and thus should not have scuttled aggregate treatment. However, if overtime involved a more complex analysis, for some reason, then the negative expected value ("NEV") interpretation of *Tyson Foods* would support the use of sampling to enable aggregation—even if Tyson had kept individual time records.

136. The class also included some former employees, and it would not be possible to videotape their current donning and doffing times since they no longer don and doff. But it would have been possible to videotape staged donning and doffing sessions by former employees. These employees might need some practice to remember how to do it, but afterward, their current times are likely to be a better approximation of their actual historical times than Mericle's sample average.

137. It is worth noting, however, that private enforcement is not the only option for employees in these cases. The FLSA also authorizes suit by the Department of Labor to recover the same monetary relief that employees can obtain. *See* 29 U.S.C. § 216(c) (2012); *Espenscheid v. DirectSat USA, LLC*, 705 F.3d 770, 776 (7th Cir. 2013) (discussing the public enforcement alternative). Even so, the FLSA clearly contemplates private enforcement, and the Supreme Court recognizes the importance of private enforcement to

2. The Underlying Cause of Action

The second of the Court's two factors—the underlying cause of action—also needs fleshing out. The elements of the cause of action define the legal and factual issues that are candidates for sampling, and the source of the cause of action—state or federal—affects whether a federal court must follow state law if sampling is substantive for *Erie* purposes.¹³⁸ These are obvious points. But the cause-of-action factor plays another important role in the analysis. The substantive policies underlying the cause of action are critical to the justification for using sampling.

The *Tyson Foods* Court makes this point clear when, quoting *Mt. Clemens*, it ties sampling to the “remedial nature of [the FLSA] and the great public policy which it embodies.”¹³⁹ The Court does refer to the FLSA as a “remedial” statute and invokes the principle that remedial statutes should be construed generously.¹⁴⁰ On a quick reading, this might lead one to conclude that sampling should be limited to remedial statutes. But this conclusion would be a mistake. There is no such thing as a special class of remedial statutes; every statute that provides remedies for legal wrongs is remedial.¹⁴¹ At most, the “remedial” label means that the policies underlying the statute *strongly* support providing relief to the favored group and thus

realizing the policy goals of the Act. See *Anderson v. Mt. Clemens Pottery Co.*, 328 U.S. 680, 687–88 (1946), *superseded by statute*, Act of May 14, 1947, ch. 52, sec. 4, 61 Stat. 84, 86–87 (codified as amended at 29 U.S.C. § 254(a) (2012)).

138. See *Erie R.R. Co. v. Tompkins*, 304 U.S. 64 (1938).

139. *Tyson Foods, Inc. v. Bouaphakeo*, 136 S. Ct. 1036, 1047 (2016) (alteration in original) (quoting *Mt. Clemens*, 328 U.S. at 687).

140. See *id.* Other courts have relied on the remedial nature of the FLSA as support for interpreting its collective action provision liberally. See, e.g., *Hoffmann-La Roche Inc. v. Sperling*, 493 U.S. 165, 173 (1989) (relying on the “broad remedial goal” of the FLSA to authorize trial judge assistance with giving notice of the collective action to employees); *Church v. Consol. Freightways, Inc.*, 137 F.R.D. 294, 306 (N.D. Cal. 1991) (holding that because the Age Discrimination in Employment Act (“ADEA”) “is a remedial statute which is to be liberally construed in light of its purpose,” ADEA plaintiffs bringing a FLSA collective action do not have to show predominance of common questions to get conditional certification and thus be able to invite other employees to join in).

141. As the Seventh Circuit observed in a case involving the Labor-Management Reporting and Disclosure Act (“LMRDA”),

Plaintiffs stress that the LMRDA is a remedial measure and seek a liberal construction. This maxim is useless in deciding concrete cases. *Every* statute is remedial in the sense that it alters the law or favors one group over another. . . . [A]fter we determine that a law favors some group, the question becomes: How *much* does it favor them?

Stomper v. Amalgamated Transit Union, Local 241, 27 F.3d 316, 320 (7th Cir. 1994).

that the statute should be construed in a way that facilitates that goal.¹⁴² But this applies to any statutory claim that promotes strong substantive policies.

It would also be a mistake to conclude from the fact that the FLSA expressly authorizes a collective action that sampling should be limited to only those statutory claims that include explicit class action, collective action, or other aggregation provisions.¹⁴³ For one thing, the fact that a statute is silent does not mean that Congress does not value enforcement through aggregation. Congress might simply have assumed that there was no need for explicit authorization because Rule 23 or some other aggregation device was already available.¹⁴⁴ More importantly, the reason that Congress empowers aggregation for statutory claims is because it believes that aggregation is important to the enforcement of the statute's substantive policies.¹⁴⁵ Thus, it is those policies that matter, not the presence or absence of congressional authorization.¹⁴⁶

This is exactly how the *Tyson Foods* Court justifies sampling—by reference to the underlying statutory policies, not the statutory text.¹⁴⁷

142. *See id.*

143. Some courts have relied on the presence of the collective action provision to justify rules that facilitate employee use of the device. *See, e.g., Hoffmann-La Roche*, 493 U.S. at 170.

144. Indeed, the reason for the FLSA's collective action provision is not just to empower aggregation but also to limit aggregation to an opt-in procedure. *See* Daniel C. Lopez, *Collective Confusion: FLSA Collective Actions, Rule 23 Class Actions, and the Rules Enabling Act*, 61 HASTINGS L.J. 275, 283–84 (2009) (summarizing the 1947 amendments to the FLSA).

145. *See, e.g., Hoffmann-La Roche*, 493 U.S. at 173 (equating Congress's decision to leave the FLSA's aggregation provision "intact" with the courts' duty to ensure that the "broad remedial goal of the statute . . . [is] enforced to the full extent of its terms"); *Watkins v. Simmons & Clark, Inc.*, 618 F.2d 398, 402 n.8 (6th Cir. 1980) (referring to the "congressional purpose" behind the Truth in Lending Act's class action provision as "using the threat of a class action to force compliance with the Act" and cautioning that subverting the class action could render "the prophylactic effect of class liability exposure . . . nonexistent"); G. W. Foster, Jr., *Jurisdiction, Rights, and Remedies for Group Wrongs Under the Fair Labor Standards Act: Special Federal Questions*, 1975 WIS. L. REV. 295, 309 (arguing that the FLSA's collective action provision was necessary to enforce the anti-unfair-competition purpose of the statute).

146. It would matter if the statute explicitly authorized, or rejected, sampling, not just aggregation. But the FLSA says nothing about sampling, and very few statutes do. For an example of a statute that expressly authorizes sampling, see 15 U.S.C. § 15d (2012) (authorizing statistical sampling to measure damages in a *parens patriae* antitrust action).

147. *Tyson Foods, Inc. v. Bouaphakeo*, 136 S. Ct. 1036, 1047 (2016) (relying on the "remedial nature" of and the "great public policy" behind the FLSA (quoting *Anderson v. Mt. Clemens Pottery Co.*, 328 U.S. 680, 687 (1946), *superseded by statute*, Act of May 14, 1947, ch. 52, sec. 4, 61 Stat. 84, 86–87 (codified as amended at 29 U.S.C. § 254(a) (2012)))).

It enlists the FLSA's substantive policies to justify sampling in an individual suit.¹⁴⁸ It then concludes that if sampling is available in an individual suit, it must also be available in a Rule 23 class action because of the REA.¹⁴⁹

Drawing on substantive law policies to justify the use of sampling makes sense in general. The primary goal of adjudication is to enforce the substantive law, and the point of enforcement is to promote the policies that the substantive law serves. Since the reason to use sampling is to better enforce those policies, the case for it must be stronger when those policies are stronger.

More precisely, in any litigation, the substantive policies at stake affect the magnitude of the expected error costs from using a procedure. Sometimes plaintiffs recover when the defendant is not in fact liable (a false positive error), and sometimes they fail to recover when the defendant is in fact liable (a false negative error).¹⁵⁰ Each of these two types of error generates its own costs, both to the losing party and to society at large (as the term is used here, "costs" include all types of negative effects, including moral as well as economic harms). Procedures often reduce one type of error while increasing the other. This is true for sampling. A sampling procedure reduces false negatives by enabling deserving plaintiffs to recover, but it also increases false positives by allowing everyone in the class to recover, including those who have suffered no legal wrong. Whether sampling is justified, therefore, depends on the relative costs of these two types of error. And those costs must be measured in terms of the substantive law values at stake.¹⁵¹

148. *See id.* It is also worth mentioning that basing FLSA liability on average times has a certain normative appeal. An average time is likely to be a reasonable time, and a reasonableness standard might seem an appropriate measure of an employer's obligation in this context. It is possible that the normative appeal of the sample average in these FLSA suits contributed to the Court's willingness to authorize its use in the case. However, the Court never mentioned this point and relying on it to justify the sample average would be tantamount to altering the FLSA, which could create separation-of-powers and REA problems. I am indebted to my colleague Larry Sager for alerting me to this point.

149. *See id.* at 1046.

150. For a more comprehensive account of error costs, see BONE, *supra* note 131, at 128–46.

151. There are two additional complications. First, expected process costs also matter. Process costs include all the costs of applying the sampling procedure, including the costs of hearings to decide whether to use it, the costs of designing it, and the costs of actually implementing it. *Id.* at 146. Second, the optimal balance of expected error and process costs depends on one's normative theory of procedure and, in particular, on whether that theory is efficiency-based or rights-based. *See* Bone, *Statistical Adjudication*, *supra* note 7,

The *Tyson Foods* Court reads the FLSA as placing a very high social value on vindicating employee rights. In effect, the Court assumes that false negatives—denying compensation to deserving employees—are much more costly than false positives—forcing employers to pay compensation to some employees whose FLSA rights have not been violated. This is the point of emphasizing the “remedial nature of [the FLSA] and the great public policy which it embodies.”¹⁵² Under these circumstances, a procedure like sampling, which reduces the more costly false negative error, is desirable as long as it does not increase false positives by too much.¹⁵³

In sum, the Court’s cause-of-action factor and its purpose factor work together to justify sampling. If the purpose of sampling is to remove a serious barrier to relief and if removing that barrier enables enforcement of important social policies served by the cause of action, without seriously impairing other relevant policies, then there is a strong argument for sampling.

3. Heterogeneity and Reliability

As we have seen, the *Tyson Foods* Court distinguishes *Wal-Mart* by relying on the degree of heterogeneity within the employee group.¹⁵⁴ Heterogeneity matters because it affects the reliability of a sampling methodology. The more that individuals within the population differ from one another, the less certain it is that the

at 595–617 (analyzing statistical adjudication under efficiency-based and rights-based theories).

152. *Tyson Foods*, 136 S. Ct. at 1047 (quoting *Anderson v. Mt. Clemens Pottery Co.*, 328 U.S. 680, 687 (1946), *superseded by statute*, Act of May 14, 1947, ch. 52, sec. 4, 61 Stat. 84, 86–87 (codified as amended at 29 U.S.C. § 254(a) (2012))).

153. As I explain below, the precise way this works is different for an efficiency theory of procedure than for a rights-based theory. *See infra* Section III.B. In an efficiency theory, for example, whether false positives are increased “by too much” depends on four variables: (1) the cost of false positives, (2) the cost of false negatives, (3) the amount by which sampling reduces the probability of false negatives, and (4) the amount by which sampling reduces expected process (litigation) costs. *See BONE, supra* note 131, at 128–55 (describing an error cost analysis). To illustrate, suppose that the cost of a false negative is 20, the cost of a false positive is 10, sampling reduces the probability of a false negative error by 10% (say from 20% to 10%), and sampling reduces expected process costs by 2. In this stylized example, the expected benefit of sampling is the savings in false negative error costs, which is 2 (i.e., $0.1 \times 20 = 2$) plus the savings in expected process costs, which is also 2—for a total benefit of 4. Therefore, if sampling increases expected false positive error costs by more than 4, costs would exceed benefits and sampling would not be desirable. Given the assumption that the cost of a false positive error is 10, sampling would increase expected false positive error costs by more than 4 if it increased the probability of a false positive error by more than 40%.

154. *See supra* notes 106–08 and accompanying text.

sample average is a good estimate of the population average or of actual values for each individual. One can handle this problem by using a larger sample size or a stratified sampling procedure (or regression); however, doing so increases litigation costs.

For a utilitarian committed to economic efficiency, what matters is the balance of costs and benefits.¹⁵⁵ As heterogeneity increases, the costs of adjusting the sampling procedure to compensate for it also increase. At some point, the marginal costs exceed the marginal benefits, and this is the point where the degree of heterogeneity bars sampling.

For someone committed to a rights-based theory, heterogeneity works a bit differently, but the conclusion is the same—it all depends on the case. For example, as we shall see, a rights-based jurist can accept sampling even when it produces a substantial divergence between average recovery and actual entitlement if the alternative is worse for rights. This was the case in *Tyson Foods* and *Wal-Mart*, where plaintiffs would not have been able to recover at all without the collective adjudication made possible through sampling.¹⁵⁶

In sum, a properly designed sampling procedure should satisfy heterogeneity constraints in many cases. This is because heterogeneity is relevant mostly as a factor in determining the reliability of the statistical methodology.¹⁵⁷

B. *Justifying Sampling's Outcome Effects*

Sampling can produce more accurate outcomes than individual trials if the heterogeneity of the aggregation is small enough and the

155. Actually, a utilitarian should be interested in the real world incentives generated by the combination of substantive and procedural law. These incentives are affected by outcome error, of course, but they can also be directly affected by procedure without regard to effects on outcome error. For example, a procedural system that produces high litigation costs might deter socially undesirable conduct in a direct way if actors anticipate being sued when they violate the law and incur high litigation costs as a result. In other words, high litigation costs are not just costs of the system; they are also benefits insofar as they add to the procedural system's deterrent effect. It follows that reducing litigation costs can weaken deterrence, all other things equal. I ignore this refinement in the text.

156. See *supra* notes 128–37 and accompanying text.

157. Reliability is usually tested in a *Daubert* hearing. See FED. R. EVID. 702; *Daubert v. Merrell Dow Pharm., Inc.*, 509 U.S. 579 (1993). It is also worth mentioning that judges might need the assistance of neutral experts, perhaps appointed as special masters or as expert witnesses, to evaluate and implement a reliable statistical approach. See FED. R. CIV. P. 53 (authorizing the appointment of special masters); FED. R. EVID. 706 (allowing the court to appoint expert witnesses).

error risk of individual suits is large enough.¹⁵⁸ When the aggregation is nearly homogeneous, the sample average is quite a good estimate of the correct result for every case. Indeed, this estimate could be even better than the trial verdict in an individual suit if that verdict is likely to deviate from the correct outcome by a large margin because of a high risk of trial error.

However, these two conditions—low heterogeneity and large individual trial error—are not easily satisfied. There is no assurance that the aggregation will have low heterogeneity, especially with respect to the individual issues for which sampling is used.¹⁵⁹ Moreover, because of limited information, judges will have great difficulty sorting between aggregations that meet the conditions and aggregations that do not.¹⁶⁰

Thus, we must address the fact that sampling can increase the error risk for at least some, and perhaps many, cases in the distribution and systematically distort outcomes relative to substantive entitlements. In *Tyson Foods*, for example, the sample average generated overtime liability for some employees who did not in fact work overtime, and it also likely denied overtime liability for some employees who did work overtime.¹⁶¹ Moreover, because everyone receives the average, sampling overcompensates deserving plaintiffs with below-average claims and undercompensates deserving plaintiffs with above-average claims.

158. See Bavli, *supra* note 5, at 82–83 (demonstrating that sampling will reduce the aggregate error risk on average if the variance of the aggregation, what Bavli calls “claim variability,” is less than the error risk for an individual suit, what Bavli calls “judgment variability”); Bone, *Statistical Adjudication*, *supra* note 7, at 578–84 (noting that whether sampling improves accuracy for all cases in the aggregation depends on the heterogeneity of the aggregation and the error risk in an individual suit); Michael J. Saks & Peter David Blanck, *Justice Improved: The Unrecognized Benefits of Aggregation and Sampling in the Trial of Mass Torts*, 44 STAN. L. REV. 815, 833–37 (1992). In considering the effect of sampling on error risk, one must be careful to distinguish between aggregate effects and case-specific effects. Sampling might reduce the expected error risk for the aggregation as a whole, but still increase and also skew the error risk for cases far out on the tail of the distribution. The aggregate effect is what matters for an efficiency theory, but, as we shall see, case-specific effects matter for a rights-based theory.

159. As noted above, sampling is useful for overcoming the predominance obstacle to certification of a Rule 23(b)(3) class action. See *supra* note 132 and accompanying text.

160. It might be possible to use sequential sampling to gather more information about the population distribution, but this is complicated and can be costly. See Bavli, *supra* note 5, at 84–85.

161. See *Tyson Foods, Inc. v. Bouaphakeo*, 136 S. Ct. 1036, 1055 (Thomas, J., dissenting). The latter cases would arise if an employee took longer than the average donning and doffing time and the difference would have pushed the employee above forty hours per week.

The seriousness of these outcome effects depends on one's normative theory of procedure. A utilitarian committed to economic efficiency cares only that outcomes generate optimal incentives overall.¹⁶² Sampling fares well by this measure.¹⁶³ As far as defendant's incentives are concerned, what matters is the defendant's total liability, and the sample average multiplied over all the cases in the aggregation is a very good estimate of that.¹⁶⁴ As for potential plaintiffs, they adjust their pre-injury behavior according to what they think will happen if they are injured. Thus, the fact that plaintiffs receive more or less than their substantive entitlements is not a problem for a utilitarian because pre-injury expectations average over all possible cases, just as the sample average does.¹⁶⁵ And even if sampling introduces some distortion, the resulting social costs might well be offset by the savings in litigation costs that sampling produces in large enough case aggregations.

Skewed outcomes are more troubling for a rights-based procedure theory. Roughly speaking, a rights-based theory assumes that parties have procedural rights that trump or substantially constrain arguments for limiting procedure based on minimizing social costs, maximizing aggregate welfare, or pursuing collective

162. See Louis Kaplow, *The Value of Accuracy in Adjudication: An Economic Analysis*, 23 J. Legal Stud. 307, 309–10 (1994) (focusing on how procedural accuracy affects *ex ante* behavior and incentives).

163. This discussion is a summary of the analysis that I develop more fully in other writing. See Bone, *Statistical Adjudication*, *supra* note 7, at 595–98.

164. More precisely, what matters for a rational and risk-neutral defendant is expected liability, and the sample average does a very good job of estimating expected liability for a representative aggregation of cases and a large enough random sample. See BONE, *supra* note 131, at 20–36 (explaining expected liability and expected value). Also, sampling reduces outcome variance and thus should improve incentives for risk-averse defendants.

165. This is true as long as the aggregation is representative of the larger population of potential cases and the sampling procedure is reliable. More precisely, a plaintiff's *ex ante* incentives depend on her expected recovery, and expected recovery is just an average over all possible future cases. To illustrate, suppose that there are 100 plaintiffs; that the defendant is not liable at all to 10 of these plaintiffs; and that among the 90 to whom it is liable, 20 suffered damages of 30,000, 50 suffered damages of 20,000, and 20 suffered damages of 10,000. If a sample of 10 cases is randomly drawn from the population of 100, we would expect 1 case of no liability, 2 cases with 30,000 damages, 5 cases with 20,000 damages, and 2 cases with 10,000 damages. Assuming perfect accuracy, if all the sample cases were tried, then the sample average would equal: $[(1 \times 0) + (2 \times 30,000) + (5 \times 20,000) + (2 \times 10,000)] \div 10 = 18,000$. Therefore, the defendant's total liability would be the sample average multiplied by the 100 cases in the population, which is $18,000 \times 100 = 1,800,000$. This is exactly the same as the defendant's total liability if all cases were tried individually, again assuming perfectly accurate trials. That total would be $(10 \times 0) + (20 \times 30,000) + (50 \times 20,000) + (20 \times 10,000) = 1,800,000$.

social goals.¹⁶⁶ The main problem from a rights-based perspective is sampling's impact on a plaintiff's rights. To be sure, the defendant's rights matter just as much as the plaintiff's. But defendant's rights are not seriously affected by a well-designed sampling procedure. This is so because defendant's total liability is roughly the same with sampling as without, and defendant's total liability is what should matter to any sensible rights-based theory.¹⁶⁷ The defendant pays some plaintiffs more than it should but it pays others less than it should, and the errors cancel out.

The effect on plaintiffs' rights is more serious because errors do not cancel on the plaintiffs' side. Some plaintiffs receive less than what their substantive rights guarantee and others receive more. Even so, the distortion is not itself disqualifying. If it were, the Court's endorsement of sampling in *Tyson Foods* would be open to criticism. More importantly, many of the situations where sampling makes the most sense are situations where skewed outcomes can be justified within a rights-based theory.¹⁶⁸

For example, if individuals are unable to sue or recover without the use of sampling, there is no reason to reject the procedure just because it distorts outcomes. As discussed above, this is the situation in *Tyson Foods*, as well as for the backpay claims in *Wal-Mart*.¹⁶⁹ In these cases, sampling makes it possible to satisfy rights claims that would otherwise go unmet. There is no sensible rights-based ground to complain about that result, at least as long as the sampling procedure is designed to minimize distortions.

Also, even if individuals can sue and recover without sampling, it matters that they save litigation costs with sampling.¹⁷⁰ If the litigation cost savings equal or exceed the shortfall due to averaging, a plaintiff will be no worse off with sampling, and might even be better off than she would in an individual suit, despite obtaining a formal judgment

166. See, e.g., RONALD DWORKIN, A MATTER OF PRINCIPLE 93–94 (1985); Robert G. Bone, *Agreeing to Fair Process: The Problem with Contractarian Theories of Procedural Fairness*, 83 B.U. L. REV. 485, 513–16 (2003).

167. Technically, the defendant's procedural right attaches to each individual suit, not to the aggregation as a whole. However, the distribution of error risk across cases is what matters from a rights-based perspective. See *infra* notes 175–78 and accompanying text. And a defendant has no ground to complain about a distribution that leaves its total liability intact.

168. See generally Lahav, *supra* note 5 (defending the use of sampling on the ground that it promotes outcome equality).

169. See *supra* notes 128–37 and accompanying text.

170. Professor Jay Tidmarsh relies on a similar point to justify his presumptive judgment approach. See Tidmarsh, *supra* note 2, at 1487–88.

or settlement less than what her right guarantees. This result should satisfy any rights-based concerns.¹⁷¹

Moreover, there is no reason to worry about outcome effects under a rights-based theory when a plaintiff consents to sampling. Indeed, a plaintiff with a high-value claim can opt out of a Rule 23(b)(3) class action and does not need to join a FLSA collective action if she believes she will fare worse with sampling.¹⁷² To be sure, plaintiffs have imperfect information when they make opt-out or opt-in decisions, and it is more difficult to infer consent when a party lacks some of the information necessary to make an informed choice. But it is important to remember that parties make lots of outcome-determinative choices in litigation with less-than-perfect information.

Finally, the essential consideration for a rights-based theory is not the magnitude of the error risk itself but the distribution of that risk across litigants and cases and, in particular, the reasons why it is distributed the way it is.¹⁷³ To see this point clearly, it is useful to begin by noting that sampling's effect in producing outcomes that

171. One might object that litigation cost savings should not be considered because they are not part of the substantive right. However, this position is untenable. If litigation-related costs are irrelevant, then it would not matter to a rights-based theory that a plaintiff who files late in the litigation queue recovers very little or nothing at all because of high delay costs or insufficient remaining assets. The formal judgment or settlement would be enough even if it was (virtually) worthless in practical terms. No litigation system that purports to respect rights can possibly be satisfied with such a result. For more on this point, see Bone, *Actuarial Litigation*, *supra* note 7, at 253–54.

172. See FED. R. CIV. P. 23(c)(2)(B)(v). Opt out might scuttle the class action if the class unravels as the sample average changes. Whether this will happen depends, among other things, on the litigation cost savings with aggregation and sampling. In any event, this is a factor to consider when deciding whether to sample, but it is not relevant to rights-based objections.

173. For an excellent discussion of this point, see DWORKIN, *supra* note 166, at 93–94; see also Robert G. Bone, *Procedure, Participation, Rights*, 90 B.U. L. REV. 1011, 1013–18 (2010). According to Professor Dworkin, procedural rights should be understood as rights to equal concern and respect in the distribution of the risk of outcome error. DWORKIN, *supra* note 166, at 92–93. Dworkin's analysis is too complicated to review in detail here. In a nutshell, he argued that outcome errors produce “moral harm” as well as “bare harm.” See *id.* at 80–81. Moral harm consists of the injustice associated with failing to honor substantive legal rights, and thus is always produced by outcome error. While bare harm is suitable for utilitarian balancing, moral harm is properly the subject of rights. Roughly, each party has a procedural right to a distribution of error risk that reflects equal concern and respect for the importance of the moral harm at stake. *Id.* at 92–93 (noting that each party has “a right to procedures justified by the correct assignment of importance to the moral harm the procedures risk, and a related right to a consistent evaluation of that harm in the procedures afforded them as compared with the procedures afforded others in different civil cases”). Thus, Dworkin focused mainly on the distribution of error risk and saw procedural rights as constraining the set of acceptable distributions.

differ from substantive entitlements cannot by itself be enough to condemn the procedure. Obviously, all procedure generates outcome error; no conceivable procedural system could possibly enforce the substantive law perfectly. So if the mere possibility of error were enough, then all procedures would have to be condemned. Moreover, it also makes no sense to specify some maximum level of error risk for all cases, because the acceptable level varies with the type of case.¹⁷⁴

For these reasons, a rights-based theory focuses on the distribution, rather than the absolute magnitude, of the error risk. It insists that this distribution treat all right holders with equal concern and respect. Whether a particular distribution satisfies this condition depends on the reasons for the distribution. A procedure like sampling can distribute error risk unevenly and still be justified provided the reasons for using the procedure accord all claimants equal concern and respect as individual right holders.¹⁷⁵ Reducing social costs is not a good enough reason because it does not respect litigants as holders of procedural rights that trump or constrain aggregate cost reduction. However, assuring fair and just treatment for all litigants can be good enough.

For example, when mass tort cases are litigated individually, plaintiffs who file late can experience serious delay if early filers clog the courts.¹⁷⁶ Delay is costly because it reduces the real value of any

174. More generally, it makes no sense to frame the procedural right as a right to a specific level of error risk. See Bone, *supra* note 166, at 513–16. There is no natural way to define the requisite level. The right cannot guarantee perfect accuracy because perfect accuracy is impossible. Nor can it guarantee maximal accuracy because there is no limit to maximal accuracy. We can always improve accuracy by trying a case multiple times and using the most frequently occurring result or the average of the trial verdicts, but that means that a right to maximal accuracy would commit society to spending all its resources on procedure—a clearly absurd result. See *id.* at 514. Finally, formulating the right as a right to a “reasonable” risk of error begs the question what is “reasonable,” and defining it as a right to a “fair hearing” begs the question what is “fair.”

175. For example, plausibility pleading, applied strictly, can make it difficult for plaintiffs to file suit when the defendant has exclusive possession of the information the plaintiff needs to make the necessary allegations, and as a result it can systematically skew error in the defendant’s favor. See Robert G. Bone, Twombly, *Pleading Rules, and the Regulation of Court Access*, 94 IOWA L. REV. 873, 925–28 (2009). This does not necessarily mean, however, that plausibility pleading violates procedural rights any more than notice pleading does (by increasing the risk of *in terrorem* settlements). It all depends on the reasons for the rule and the resulting error-risk distribution. See *id.* at 912–15 (describing a rights-based analysis of pleading rules).

176. See, e.g., *Ortiz v. Fibreboard Corp.*, 527 U.S. 815, 866–67 (1999) (Breyer, J., dissenting) (noting the extremely long delays in adjudicating asbestos cases due to the huge number of filed lawsuits, and the resulting impact on plaintiffs). Also, in a limited fund case, where the defendant does not have enough assets to satisfy all the plaintiffs’

recovery. The order of filing, however, is often just a matter of luck, contingent on when an injury happens to occur or become manifest, and there is no moral reason why luck should make a difference to what a plaintiff recovers. Aggregation reduces delay cost and evens out its impact,¹⁷⁷ and sampling makes aggregation possible. It is true that plaintiffs with high-value claims receive less than their substantive entitlements.¹⁷⁸ But not everyone can recover fully. The choice, therefore, is between giving full recovery to some and nothing to others, or distributing the error risk in a way that treats each person as an equal right holder. A rights-based theory opts for the latter and justifies sampling as a means to that end.

In sum, it is important that a court justify the use of sampling in a way that addresses systematic outcome effects. Sampling is relatively easy to justify within an efficiency-based theory, but a rights-based theory imposes tighter constraints. Still, no matter which theory one adopts, sampling can always be justified when it makes possible the litigation of small claims, when its litigation cost savings compensate for any outcome shortfall, or when plaintiffs consent to its use. It can also be justified when the aggregation it supports improves the justness and fairness of the error risk distribution.

C. Other Factors

There are additional factors that should be considered when deciding whether to use sampling. I have discussed many of these in my other writing,¹⁷⁹ and I will only summarize them here.

First, by speeding up recovery and reducing litigation and delay costs, sampling attracts more lawsuits, and the resulting increase in litigation can produce excessive deterrence.¹⁸⁰ This is especially

claims, those plaintiffs who file suit early obtain full recovery while those who file late, after the defendant's limited assets have been exhausted, receive nothing beyond a formal judgment. See 7AA CHARLES ALAN WRIGHT, ARTHUR R. MILLER & MARY KAY KANE, FEDERAL PRACTICE AND PROCEDURE § 1774, at 30–32 (3d ed. 2005) (citing Norman C. Sabbey, *Rule 23: Categories of Subsection (b)*, 10 B.C. INDUS. & COM. L. REV. 539, 541 (1969)). Aggregation through sampling helps redress this inequality by permitting an equitable distribution that binds everyone. See *id.* § 1774, at 32.

177. See *Ortiz*, 527 U.S. at 866–68 (Breyer, J., dissenting); *Amchem Prods., Inc. v. Windsor*, 521 U.S. 591, 632–33 (1997) (Breyer, J., concurring in part and dissenting in part).

178. A trial judge can, and in some cases should, reduce the divergence by using a stratified sample or possibly even regression techniques.

179. See Bone, *Actuarial Litigation*, *supra* note 7, at 239–59; Bone, *Statistical Adjudication*, *supra* note 7, at 576–94.

180. Bone, *Statistical Adjudication*, *supra* note 7, at 596.

problematic when the aggregation made possible by sampling increases liability exposure beyond what Congress envisioned when it created the substantive claim.¹⁸¹ Judges should be aware of these risks and possibly deny sampling when they are too serious. But the possibility of overdeterrence in some cases should not bar sampling across the board.

Second, sampling can reduce the amount plaintiffs are willing to invest in litigation relative to the defendant, and the resulting asymmetry can skew error risk in the defendant's favor.¹⁸² However, there are many factors that affect litigation investment incentives, and it is not at all clear how significant this particular one is. In any event, as I have explained elsewhere, the asymmetry can be reduced by careful choice of the sampling protocol.¹⁸³

Third, sampling can reduce incentives to settle by affecting the size of the settlement surplus.¹⁸⁴ This is not necessarily a serious concern, however, because many factors affect settlement incentives. Moreover, the problem with reduced settlement incentives is that it produces more trials, and sampling might save enough in litigation costs to compensate for any trial costs it adds.

Fourth, any evaluation of sampling must take account of its impact on the filing of frivolous claims.¹⁸⁵ Since sampling relieves

181. Although they do not involve sampling, class actions for statutory damages can impose crippling liability beyond what Congress likely intended, especially in cases involving minor, if any, actual harm. *See, e.g., Parker v. Time Warner Entm't Co.*, 331 F.3d 13, 21–22 (2d Cir. 2003) (recognizing the risk that class certification could lead to a potentially devastating damages award disproportionate to the harm actually suffered in a case with twelve million individuals seeking statutory damages authorized by the Cable Communications Policy Act).

182. *See Bone, Statistical Adjudication, supra* note 7, at 589–90. For example, suppose a court has an aggregation of 1,000 cases and tries a sample of 100. The defendant will invest in the 100 sampled cases based on its expected liability over all 1,000 cases. If sample-case plaintiffs receive the sample average, however, they will invest suboptimally because they have incentives to free ride on other sample-case plaintiffs and are not able to capture all the benefits they confer on the other plaintiffs in the aggregation. Thus, free riding and externalities can lead plaintiffs to invest less in the sampled cases than the defendant, which should increase the defendant's chance of winning assuming that a party is more likely to win the more she invests relative to her opponent.

183. *Id.* at 590–91. The asymmetry can be minimized in many cases by giving the sample average to all the plaintiffs in the aggregation and spreading the litigation costs evenly among them. *See id.* As it happens, these conditions were satisfied in the *Tyson Foods* case, where the sample average was applied to all employees in the class and sampling costs were borne proportionately by all class members. *See Tyson Foods, Inc. v. Bouaphakeo*, 136 S. Ct. 1036, 1043–44 (2016).

184. *See Bone, Actuarial Litigation, supra* note 7, at 244–47.

185. *Id.* at 250–51.

litigants of the need to prove liability and damages on an individualized basis, it creates incentives for lawyers to pack the class or other aggregation with frivolous and weak claims. However, this adverse selection problem inheres in all aggregations, whether or not they use sampling. For example, class actions frequently end in aggregate settlements based on average claim values, and this averaging process creates incentives to include as many claims as possible in the class, regardless of merit. Yet this risk is not thought serious enough to reject the class action device outright. Indeed, the class in *Tyson Foods* included some employees who had meritless claims, but this did not stop the Court from allowing sampling. The risk of frivolous filings should be an important factor in judicial decisions whether to authorize sampling, just as it should be in judicial decisions whether to certify a class or approve a class settlement. But sampling should not be denied just because it encourages the filing of some frivolous claims.

Fifth, depending on how it is structured, sampling can deny participation opportunities to many plaintiffs.¹⁸⁶ When everyone receives the sample average, only those plaintiffs in the sampled cases have a chance to litigate the issues.¹⁸⁷ The significance of this fact, however, depends on whether individual participation is valued for outcome quality reasons or process-oriented dignitary reasons.¹⁸⁸ If outcome quality is the focus, there is no cause for concern as long as the sample plaintiffs vigorously litigate their cases and the sample is representative of the aggregation. When a case is vigorously litigated, there is no reason to believe that a second case will produce a more accurate result from a social perspective. The second plaintiff might win when the first loses, but that does not mean that the result in the second case is more accurate than the result in the first.

Dignity is a different matter. Many proceduralists believe that personal participation and control are required to respect the dignity of persons affected by litigation, entirely apart from any impact on

186. For an in-depth discussion of this point, see Bone, *Statistical Adjudication*, *supra* note 7, at 617–50.

187. *See id.* at 618. The Court might allow non-sampled plaintiffs to intervene in the sampled cases, but managing a large set of plaintiffs is likely to be very burdensome and will end up diluting individual participation if the judge has to appoint a litigation steering committee.

188. *See* JERRY L. MASHAW, *DUE PROCESS IN THE ADMINISTRATIVE STATE* 158–253 (1985).

outcome quality.¹⁸⁹ Even the United States Supreme Court recognizes a robust due process right to a personal “day in court” that is difficult to justify except with a dignitary theory.¹⁹⁰

A dignity-based participation right, however, is not an insurmountable barrier to sampling. For one thing, the class action already curtails participation opportunities, and this is considered justifiable because the named plaintiffs represent the interests of absent class members.¹⁹¹ By this same logic, it should be possible to justify sampling as part of that representational relationship—the named plaintiffs and class attorney choose sampling on behalf of the class, just as they choose any other litigation strategy.

More generally, the participation right is best understood as a flexible institutional right, not an absolute autonomy right.¹⁹² By

189. See, e.g., *id.* at 162–63; Martin H. Redish & William J. Katt, *Taylor v. Sturgell, Procedural Due Process, and the Day-in-Court Ideal: Resolving the Virtual Representation Dilemma*, 84 NOTRE DAME L. REV. 1877, 1893–94 (2009); Lawrence B. Solum, *Procedural Justice*, 78 S. CAL. L. REV. 181, 275–77 (2004) (linking individual participation to legitimacy); see also Elizabeth Chamblee Burch, *Litigating Together: Social, Moral, and Legal Obligations*, 91 B.U. L. REV. 87, 103–11 (2011) (exploring the non-outcome-based, collectively-oriented benefits of aggregation and implications for procedure).

190. See *Taylor v. Sturgell*, 553 U.S. 880, 892–93 (2008) (quoting *Richards v. Jefferson Cty.*, 517 U.S. 793, 798 (1996)); Robert G. Bone, *The Puzzling Idea of Adjudicative Representation: Lessons for Aggregate Litigation and Class Actions*, 79 GEO. WASH. L. REV. 577, 585–86 (2011) (explaining why *Taylor*’s strong day-in-court right can be supported only by a process-oriented participation theory).

191. *Taylor*, 553 U.S. at 894 (recognizing the class action as an exception to the general rule against nonparty preclusion).

192. Bone, *supra* note 190, at 614–16. It might be tempting to invoke the balancing test of *Mathews v. Eldridge*, 424 U.S. 319 (1976), to justify limits on the day-in-court participation right and make room for sampling. See *id.* at 334–35; see also *Hilao v. Estate of Marcos*, 103 F.3d 767, 786–87 (9th Cir. 1996) (applying the *Mathews* test to address a due process challenge to sampling). There are at least three problems with this strategy. First, the *Mathews* test focuses mainly on the instrumental value of procedure in producing accurate outcomes, whereas the day-in-court right focuses on procedure’s intrinsic value in respecting individual dignity. See Robert G. Bone, *Rethinking the “Day in Court” Ideal and Nonparty Preclusion*, 67 N.Y.U. L. REV. 193, 202–03, 264–79 (1992) (arguing that the Supreme Court’s broad right to a personal day in court, strictly applied, best fits a process-oriented dignitary theory of participation). Second, insofar as the *Mathews* test is utilitarian, it is incompatible with participation as a right. See Jerry L. Mashaw, *The Supreme Court’s Due Process Calculus for Administrative Adjudication in Mathews v. Eldridge: Three Factors in Search of a Theory of Value*, 44 U. CHI. L. REV. 28, 47–49 (1976) (arguing that the *Mathews* test is utilitarian). Third, and most importantly, the *Mathews* test does not fit the Supreme Court’s day-in-court jurisprudence. The Court has never invoked the *Mathews* balancing test to analyze due process challenges to nonparty preclusion; instead it relies on a supposed “deep-rooted historical tradition that everyone should have his own day in court.” *Taylor*, 553 U.S. at 892–93. Moreover, the Court recognizes an extremely broad day-in-court right that guarantees expansive individual control over litigation choices, a degree of control that exceeds anything the *Mathews* test

calling it an institutional right, I mean that the participation right is defined by the balance of considerations relevant to assuring that adjudication works fairly and justly for all litigants. The fact that the participation right functions as a *right* necessarily means that it excludes utilitarian reasons for limiting participation based on social welfare maximization. But the fact that it functions as an *institutional* right means that it tolerates limits that substantially further the institutional goals of adjudication. These include limits that make meaningful recovery possible for all claimants and even limits that reduce litigation costs when those costs are so high that they interfere with the effective functioning of the institution.¹⁹³

Sixth, any use of sampling must take account of Rules Enabling Act (“REA”) constraints. The REA authorizes the Supreme Court to make rules of practice and procedure, including the Federal Rules of Civil Procedure.¹⁹⁴ The REA includes a proviso that no Rule can “abridge, enlarge, or modify any substantive right.”¹⁹⁵ The impact of the REA proviso on the legitimacy of sampling is examined in Part IV.¹⁹⁶ Here it is useful to focus on a particular sampling constraint that the *Tyson Foods* Court derives from the REA: sampling is permissible in a class action *only* when it is also permissible in an individual suit.¹⁹⁷

This is an unfortunate interpretation of the REA. It forces a judge contemplating the use of sampling in a class action to first justify its use in individual suits. This might not be too difficult if the same reasons apply to both class and individual litigation. But it makes no sense to require the extra step when sampling is appropriate precisely because cases are aggregated. The class action,

would support. *See id.* at 898–901. As I argue in the text, the participation right is limited, but its limits do not derive from the *Mathews* balancing test. Rather, they derive from the fact that participation is an institutional right subject to institutional constraints that still respect its status as a right.

193. The right to jury trial must be considered as well. This is a complicated issue that I will not explore here. Whether sampling creates a jury trial issue depends on the sampling protocol. In particular, there is no problem if sampling is used to derive information about averages that is then presented to a jury, as in *Tyson Foods*. *See Tyson Foods, Inc. v. Bouaphakeo*, 136 S. Ct. 1036, 1043–44 (2016).

194. 28 U.S.C. § 2072(a) (2012).

195. *Id.* § 2072(b).

196. *See infra* Section IV.B.2.

197. *See Tyson Foods*, 136 S. Ct. at 1048 (distinguishing *Wal-Mart* by arguing that the intra-class heterogeneity in *Wal-Mart* precluded use of sampling in individual suits and that because of that, “[p]ermitting the use of that sample in a class action . . . would have violated the Rules Enabling Act by giving plaintiffs and defendants different rights in a class proceeding than they could have asserted in an individual action”).

for example, gathers all cases within a single court's jurisdiction and thus makes it easier to draw a random sample.¹⁹⁸ Also, the most compelling reasons to use sampling often focus on socially valuable aggregations and on sampling's contribution in facilitating them. If the only sensible interpretation of the REA required the *Tyson Foods* limitation, then we would have to accept it as a misguided constraint imposed by Congress. But, as explained in Part IV, this limitation is not the only sensible interpretation. When sampling is used to remove obstacles to the efficient, fair, and just enforcement of substantive rights, it serves the same purposes as any other procedural rule, and there is no reason why the REA should bar it.

IV. THE LEGITIMACY PROBLEM

The analysis in Part III showed that sampling, when properly designed, can be the best, and sometimes the only, way to enable private enforcement of the substantive law and assure fair and just outcomes for all injured parties. It also showed that concerns about heterogeneity and day-in-court participation rights, while relevant, do not prevent the use of sampling in many cases where it otherwise makes sense. Even so, it is likely that many readers will still have nagging doubts. For them, I suspect, a process of deciding cases by statistically extrapolating from a random sample just does not seem legitimate for adjudication. But it is not easy to explain why.¹⁹⁹

The following discussion first distinguishes between perceived legitimacy and normative legitimacy and argues that any convincing critique of statistical adjudication must be cast in normative terms. It then examines three normative arguments against the use of sampling. The first is that sampling is such a radical departure from ordinary adjudication that it risks jeopardizing the integrity of the institution. The second is that sampling's effects make it too substantive to qualify as a legitimate procedural device suitable for use without legislative approval. The third focuses on the way

198. If individual suits were litigated separately, all the judges involved in the various lawsuits would have to know about the other pending cases, agree on the use of sampling, and coordinate its implementation. In addition, if sample cases are tried and the average (or other statistic) is used for all the rest, then the court might be required to have personal jurisdiction over the sampled plaintiffs and possibly venue over their suits as well.

199. There are other situations where intuitions about legitimacy are strong but not easy to justify. For example, I expect most people would think it illegitimate to conduct an election for President of the United States by randomly sampling the electorate and basing the outcome on the sample results—even if a reliable sampling procedure could be designed and sampling was much less expensive than allowing everyone to vote.

sampling alters the decision-making process: it substitutes a mechanical algorithm for case-specific reasoning.

A. Normative Legitimacy Versus Perceived Legitimacy

There are two different types of legitimacy: perceived legitimacy and normative legitimacy. Perceived legitimacy focuses on subjective perceptions of legitimacy and worries that the public might lose faith in the court system.²⁰⁰ Normative legitimacy focuses on whether the procedure in question is legitimate on normative grounds, independent of public perceptions.²⁰¹

There are several problems with using perceived legitimacy to critique statistical adjudication. First, it seems far-fetched to think that the public will lose faith in the court system if sampling is used selectively.²⁰² In fact, the public might respond favorably if sampling speeds recovery and helps to compensate deserving victims. Second, public perceptions are circular. The more sampling is used, the more familiar the public will become with it and the more legitimate it will seem. Third, and most important, criticisms based on perceived legitimacy are often normative legitimacy critiques in disguise. A critic might believe that the particular practice is normatively illegitimate, assume that everyone else must share the same belief, and conclude that the public will reject the practice as illegitimate. By framing a normative intuition as an empirical claim about public perceptions, the critic never has to explain the theoretical basis for her intuition. For these reasons, any legitimacy critique of statistical adjudication should be framed in normative terms.

B. Three Arguments from Normative Legitimacy

This Section critically examines three possible critiques based on normative legitimacy: (1) that sampling is too strange a fit with adjudication, (2) that it is too substantive to qualify as procedure, and (3) that it is too mechanical to substitute for the ordinary judicial reasoning process. Before proceeding, however, it is important to address two other arguments that are often presented as legitimacy critiques.

200. See Bone, *supra* note 192, at 233–35.

201. *Id.* at 236.

202. For example, I am not aware of any evidence of public opposition to the limited use of bellwether trials and statistical adjudication today.

The first argument simply treats the negative effect on participation rights as a matter of institutional legitimacy, not just individual dignity.²⁰³ This argument fares no better cast in legitimacy terms than it does cast in dignity terms. All the same responses apply with equal force.²⁰⁴

The second legitimacy argument invokes the dichotomy, common in procedure literature, between two models of litigation: the “traditional model” and the “public law model.”²⁰⁵ The traditional model supposes that the primary function of civil adjudication is to resolve private disputes.²⁰⁶ The public law model envisions the purpose of adjudication more broadly, as involving regulation and public-norm creation and enforcement.²⁰⁷ Relying on this dichotomy, a critic might argue that sampling is not legitimate because it departs too much from the traditional model.

This argument, however, goes absolutely nowhere. Civil adjudication has never been merely about resolving private disputes. Insofar as it resolves disputes, it does so *according to the substantive law*.²⁰⁸ The purpose is to enforce substantive law rights, and the point of doing that is to promote the values the substantive law embodies. Thus, adjudication at its core is regulatory. Given this, a critic of sampling must explain why it is not legitimate to use sampling to improve substantive law enforcement. Put differently, the fact that sampling diverges from the traditional model means nothing because

203. See, e.g., MARTIN H. REDISH, *WHOLESALE JUSTICE: CONSTITUTIONAL DEMOCRACY AND THE PROBLEM OF THE CLASS ACTION LAWSUIT* 135–37 (2009); Solum, *supra* note 189, at 279–80.

204. See *supra* notes 186–93 and accompanying text.

205. Professor Abram Chayes, in his famous 1976 article, posited two models of litigation: the traditional model and the public law model. See Abram Chayes, *The Role of the Judge in Public Law Litigation*, 89 HARV. L. REV. 1281, 1282–84 (1976). And others have framed a similar dichotomy using different terminology. See, e.g., Meir Dan-Cohen, *Bureaucratic Organizations and the Theory of Adjudication*, 85 COLUM. L. REV. 1, 1–7 (1985) (highlighting differences between the “arbitration model” and the “regulation model”); Owen M. Fiss, *The Social and Political Foundations of Adjudication*, 6 LAW & HUM. BEHAV. 121, 122–25 (1982) (parsing the “dispute resolution model” and the “structural reform” model); Kenneth E. Scott, *Two Models of the Civil Process*, 27 STAN. L. REV. 937, 937–39 (1975) (distinguishing the “Conflict Resolution Model” and the “Behavior Modification Model”).

206. See, e.g., Chayes, *supra* note 205, at 1282–84.

207. See *id.* at 1282–84, 1302.

208. Indeed, this is why the system has motions to dismiss, broad discovery, extensive pretrial and trial procedures, and other elaborate features aimed at preventing strategic abuse and generating reasonably accurate outcomes.

the traditional model is not an accurate depiction of our litigation system.²⁰⁹

1. Sampling Is Too Strange

Statistical adjudication, at first glance, seems quite foreign to civil adjudication, as it is customarily practiced. Courts do not decide cases by averaging over a sample; they engage the facts of each case and reason to the best decision based on those facts. Still, the oddity of sampling is not a problem in itself. The merger of law and equity must have seemed odd to mid-nineteenth century American judges and lawyers accustomed to a dual system, and the adoption of liberal pleading and broad discovery in the 1938 Federal Rules of Civil Procedure must have seemed strange to jurists at that time. Still, all these innovations were eventually accepted as established features of the litigation landscape.²¹⁰

The way to make sense of this criticism is to treat it as an argument based on a Burkean conservatism about institutions and the risks of radical reform.²¹¹ There is some wisdom in proceeding cautiously when altering an extremely complex institution like civil adjudication, especially an institution that has evolved over centuries. A reform that might seem sensible can produce unintended consequences that change the institution in irreversible ways.

There are, however, two problems with this argument as a critique of sampling. First, sampling—and statistical decision making

209. For a more thorough discussion of the problems with the traditional/public-law dichotomy, see Robert G. Bone, *Lon Fuller's Theory of Adjudication and the False Dichotomy Between Dispute Resolution and Public Law Models of Litigation*, 75 B.U. L. REV. 1273, 1275–76 (1995). In addition to these two concerns, some people might worry that sampling introduces too much mathematics or smacks too much of law-and-economics. But neither of these is a serious objection. In fact, the tools of mathematics and law-and-economics have much to offer procedure. Some people might worry that sampling will dehumanize the litigation process over the long run, but this is an unrealistic fear if judges use sampling selectively and only when it is well justified.

210. Pleading standards and broad discovery are targets of criticism, of course, but few argue that they are illegitimate.

211. One might frame this argument in terms of tradition, but then one must explain why tradition should be the touchstone of legitimacy. Indeed, if legitimacy required following traditional practices, much of contemporary civil procedure would have to be jettisoned, since much of it represents a major departure from traditional practices. See FED. R. CIV. P. 19, 24 advisory committee's notes to 1966 amendment (describing the way that the 1966 reforms of Rules 19 and 24 departed from previous practice); Robert G. Bone, *Mapping the Boundaries of a Dispute: Conceptions of Ideal Lawsuit Structure from the Field Code to the Federal Rules*, 89 COLUM. L. REV. 1, 104–07 (1989) (describing how the 1938 Federal Rules of Civil Procedure altered traditional joinder rules).

more generally—are not all that strange to civil adjudication. In fact, as explained below, they are already well-entrenched features of adjudication, just not in obvious ways. Second, sampling is too sensible a tool to block its use on this ground. More generally, Burkean conservatism should not paralyze sensible reforms, especially reforms that target rapidly changing conditions.

a. Sampling Is Not That Strange After All

Sampling and decision making based on statistical averaging are not that strange in adjudication. When a lawmaker fashions a general rule of substantive law, she designs the rule with an eye to what is typical or average for the type of scenario she is trying to regulate.²¹² Moreover, to determine what is typical, the lawmaker focuses on what she knows, which is necessarily a subset or sample of all the possible scenarios—past, present, and future. It follows that when the rule is applied to a particular case, the resulting decision is based indirectly on the lawmaker's sample.

Consider a simple example. Suppose lawmakers adopt a law that prohibits drivers from exceeding fifty miles per hour on a stretch of highway. This law is based, among other things, on predictions and statistics about the risk of accidents at different speeds. Those predictions, in turn, must average over a sample of driving experience. It follows that when a judge finds a defendant liable for driving sixty miles per hour, the judge engages in a kind of statistical adjudication. Driving sixty miles per hour might have been perfectly safe given the weather, traffic, and road conditions that the defendant faced at the time. But the defendant's liability depends not on his own driving conditions, but instead on the average case that the rule was designed to target, which itself is based on a sample of driving experience. To be sure, whether the defendant exceeded fifty miles per hour is a case-specific determination, but whether the defendant's speed was unsafe is not. This is, after all, how general rules work.

Settlements also rely on sampling.²¹³ Parties value their cases for settlement purposes based on results in similar cases. In particular, each side estimates the likelihood of trial success, the probable trial

212. See generally Louis Kaplow, *Rules Versus Standards: An Economic Analysis*, 42 DUKE L.J. 557 (1992) (discussing the difference between rules and standards).

213. This is particularly significant because only about two percent of filed civil cases are ever tried; most settle and the rest are dismissed or dropped. See Marc Galanter, *The Vanishing Trial: An Examination of Trials and Related Matters in Federal and State Courts*, 1 J. EMPIRICAL LEGAL STUD. 459, 459, 462–63 tbl.1 (2004).

award, and the costs of litigating by averaging over a sample of previous cases.²¹⁴ Thus, if the case settles, the plaintiff receives an average recovery based on the sample, which is the same as she would receive through statistical adjudication.²¹⁵

Finally, jury verdicts are also based on assumptions about average behavior derived from experience. This is true no matter what view of jury decision-making one holds.²¹⁶ In Bayesian theory, jurors use conditional probabilities to update their prior beliefs that a particular hypothesis is true (such as that the defendant was drunk while driving).²¹⁷ Conditional probabilities are based on average expectations over a juror's slice of real world experience.²¹⁸ If jurors

214. See BONE, *supra* note 131, at 71–78 (describing the standard model of settlement). Suppose that defendant's lawyer is aware of 100 similar cases that were tried previously. Suppose that 40 of these cases ended in plaintiff victories and 60 ended in defendant victories. Also, of the 40 plaintiff victories, 10 cases ended in trial verdicts of \$200,000, 20 cases in verdicts of \$150,000, and 10 cases in verdicts of \$100,000. Finally, in 50 of the 100 cases, it cost defendants \$25,000 to litigate the case through trial, and in the other 50, it cost only \$15,000. Defendant's lawyer should estimate a 40% chance of plaintiff winning at trial, which is the average of total past wins and losses: $40 \text{ wins} \div 100 \text{ cases} = 40\%$. She should also estimate an expected trial verdict of \$150,000, which is the average of all the verdict amounts in the 40 cases that plaintiffs won: $[(10 \times 200,000) + (20 \times 150,000) + (10 \times 100,000)] \div 40 = 6,000,000 \div 40 = \$150,000$. And she should estimate expected litigation costs of \$20,000, the average of the litigation costs over the 100 cases: $[(50 \times 25,000) + (50 \times 15,000)] \div 100 = \$20,000$. Thus, defendant's lawyer will calculate an expected trial loss of $(0.4 \times 150,000) + 20,000 = \$80,000$. As a result, a risk-neutral defendant should be willing to pay up to \$80,000 to settle the case. The important point to note here is that all these estimates are averages over a sample of 100 cases.

215. Indeed, random sampling through statistical adjudication can produce a more accurate average value for claims than informal settlements distorted by informational asymmetries. One might wonder about the propriety of comparing sampling for trial with settlement, but it is a perfectly sensible comparison. Settlements are outcomes of the litigation process just as much as trial judgments are. As such, settlements must be evaluated by the same substantive law metric that applies to trial judgments. It is true that settlements are supported by consent, but parties can also consent to statistical adjudication. To be sure, consent to statistical adjudication usually comes from the attorney and is susceptible to agency problems, but agency problems also infect settlements. Because unsophisticated clients depend on their attorneys for advice, an attorney can nudge the client towards a settlement that serves the attorney's interests more than the client's. In addition, consent to a settlement is normatively problematic when the litigation alternative is infected by high delay or transaction costs. My point is only that consent does not mark a clear distinction between settlements and sampling.

216. See generally Michael S. Pardo, *The Nature and Purpose of Evidence Theory*, 66 VAND. L. REV. 547 (2013) (describing two alternative conceptions of evidence and proof—probabilistic versus explanatory).

217. For a brief description of Bayesian theory, see *id.* at 575 n.112.

218. See, e.g., Ronald J. Allen & Michael S. Pardo, *The Problematic Value of Mathematical Models of Evidence*, 36 J. LEGAL STUD. 107, 108 (2007). More precisely,

instead engage in a more holistic process of “inference to the best explanation,” each juror will assess the “relative plausibility of the competing hypotheses advanced by the parties”²¹⁹ by comparing these hypotheses to her own “stories, scripts, and scenarios.”²²⁰ Those stores, scripts, and scenarios are composites of a juror’s experience, which is just an informal sample of reality. Moreover, the most salient stories, scripts, and scenarios are likely to homogenize or average over that informal sample.²²¹

jurors are assumed to rely on a likelihood ratio, which is the ratio of the probability of discovering and receiving particular evidence if the hypothesis is true divided by the probability of discovering and receiving the same evidence if some other hypothesis is true. See Pardo, *supra* note 216, at 575 n.112. In these models of evidentiary inference, the likelihood ratio measures the probative value of the evidence. See Allen & Pardo, *supra*, at 108. Bayes’s Theorem gives a way to update one’s assessment of the probability a hypothesis is true by combining one’s prior probability with the likelihood ratio for a particular piece of evidence.

219. Allen & Pardo, *supra* note 218, at 136.

220. See Ronald J. Allen, *Factual Ambiguity and a Theory of Evidence*, 88 NW. U. L. REV. 604, 629 (1994); Michael S. Pardo & Ronald J. Allen, *Juridical Proof and the Best Explanation*, 27 LAW & PHIL. 223, 234–35 (2008). Professor Allen describes the process this way:

[W]e carry around stories, scripts, and scenarios, all more or less robust and all subject to modification in a virtually infinite variety of ways. “Evidence” in the human brain is embedded in these contexts, although fluidly in a way that permits constant cycling back and forth between the data, new and old, and their stories. It is into this bubbling cauldron that the trial “evidence” must enter, and its effect is determined by the interaction of the data with the prior knowledge stored in the related chunks constituting stories. Trial observations instantiate these stories and work modifications of them. This, of course, merely reemphasizes the dynamic nature of evidence and highlights that the data at trial that the conventional theory refers to as evidence is merely the means by which the preexisting stories and scenarios of factfinders are put into play, as it were, in order to fashion an estimate of what actually happened. . . . The structure of proof requires selection over the stories advanced at trial, and for data to be coherent it must be embedded, or be able to be embedded, in stories highly analogous if not identical to the stories being advanced by the parties as their claims about what happened.

Allen, *supra*, at 629–30. It is worth mentioning that inference to the best explanation is not logically inconsistent with Bayesian updating, but the type of Bayesian updating that would be required is far too complex to make it feasible. *Id.* at 607.

221. Of course, not all evidence is equally vulnerable to this averaging process. Direct evidence, for example, can provide very strong support for the ultimate fact. But even direct evidence must be evaluated for credibility and reliability, which involves a juror’s experience and her stories, scripts, and scenarios. In the *Tyson Foods* case, for example, a video showing donning and doffing by a specific employee or an average for that employee over multiple donning and doffing activities would be extremely strong evidence of donning and doffing times for that employee. Yet jurors must still assess the evidence in light of the possibility that the data is defective or the sampling methodology flawed, and

Thus, while we tend to think of adjudication as providing an individualized outcome for each particular case, adjudicative outcomes, to a large extent, average over samples of similar cases. This means that we should revisit our initial reaction to sampling. To be sure, it is unusual for a judge to decide a case by relying directly on a statistic derived from a sample. But it is not clear that it is sufficiently unusual to justify rejecting the procedure outright.

b. The Value of Sensible Reform

The litigation system as we know it today took shape through active reform efforts that responded to changing economic, social, political, and intellectual conditions.²²² The same is true for statistical adjudication. It too responds to changing conditions—in this case, the creation of new causes of action, the advent of mass harms, improvements in technology, and so on. Reforms like these are particularly useful when change is too rapid for the gradual process of institutional evolution to respond effectively.

Still, it is important to guard against overconfidence and excessive reform zeal. There are risks to altering any complex system, and one must account for the risks in advance whenever possible. Moreover, institutions have core elements that are essential to their proper functioning, and one must be careful not to jeopardize those elements. Below I argue that one such element for civil adjudication, perhaps the most important, is the reasoning process judges use to decide cases. But I also argue that statistical adjudication does not endanger that process as long as it is used properly. In the end, sampling is too sensible and useful a tool in the current world of litigation for a Burkean concern to prevent its use.

2. Sampling Is Too Substantive

One might worry that sampling should be implemented as substantive law, not as procedure, because of its outcome effects. The concern is the same as the one discussed above: sampling systematically and predictably skews individual recoveries away from substantive entitlements. This aspect raises potential problems under the Rules Enabling Act (“REA”), with its proviso that court-made

to do that, each juror must apply his or her experience—stories, scripts, and scenarios—to assess competing expert testimony.

222. See Bone, *supra* note 211, at 9–12, 78–80 (discussing the reform efforts that led to the field code and those that led to the Federal Rules of Civil Procedure).

rules must not “abridge, enlarge, or modify any substantive right”²²³ and also implicates separation-of-powers concerns more generally. The argument is that if sampling alters substantive rights by skewing outcomes, it can be used legitimately only when it is authorized by substantive rules made in the usual way substantive law is made.²²⁴

This is the concern that animated the *Wal-Mart* Court’s rejection of statistical adjudication for backpay awards.²²⁵ While a bit opaque, the Court’s argument is fairly easy to understand. Because the Ninth Circuit’s proposed sampling procedure would have afforded relief to some class members who had no right to it (because Wal-Mart had a good defense that it was not allowed to litigate), that procedure “enlarged” the substantive rights of those class members in violation of the REA.²²⁶ The problem is that this reading of the REA threatens all forms of statistical adjudication. At its core, statistical adjudication depends on sampling, and sampling works by cutting off the individual litigation of rights and defenses for cases that are not part of the sample. To be sure, the REA applies only to those procedural rules, like the Federal Rules of Civil Procedure, enacted pursuant to the formal court rulemaking process.²²⁷ But separation-of-powers principles extend more broadly and create similar issues for sampling implemented through judge-made common law.

Fortunately for the future of statistical adjudication, *Wal-Mart*’s interpretation of the REA is not the Court’s last word on the subject. *Tyson Foods* reads *Wal-Mart* narrowly to create some room for sampling.²²⁸ According to the *Tyson Foods* Court, sampling is

223. 28 U.S.C. § 2072(b) (2012); *see supra* notes 194–95 and accompanying text.

224. *See, e.g., In re Chevron U.S.A., Inc.*, 109 F.3d 1016, 1023 (5th Cir. 1997) (Jones, J., concurring) (observing that “there is a fine line between deriving results from trials based on statistical sampling and pure legislation” and that the court is “not authorized by the Constitution or statutes to legislate solutions to cases in pursuit of efficiency and expeditiousness”); *cf. In re Fibreboard Corp.*, 893 F.2d 706, 712 (5th Cir. 1990) (“We are told . . . that the difficulties faced by the courts as well as the rights of the class members to have their cases tried cry powerfully for innovation and judicial creativity. The arguments are compelling, but they are better addressed to the representative branches—Congress and the State Legislature.”).

225. *See Wal-Mart Stores, Inc. v. Dukes*, 564 U.S. 338, 367 (2011).

226. *See id.* Although the Court referred to the defendant’s “right to litigate” defenses, it seems unlikely that the Court meant to focus on the *process* of litigating. It is much more likely that the Court meant to focus on the *outcome* of litigating when the defendant is unable to present its defenses.

227. 28 U.S.C. § 2072.

228. *See Tyson Foods, Inc. v. Bouaphakeo*, 136 S. Ct. 1036, 1048 (2016). The *Tyson Foods* Court’s reading of *Wal-Mart* is highly questionable. It supposes that the *Wal-Mart* Court used the likely result in an individual suit as the REA baseline. *See id.* However, a

permitted in a class action if—and only if—it is also permitted in individual suits.²²⁹ As discussed above, this limitation is undesirable as a policy matter because it excludes cases where sampling makes sense only in an aggregation.²³⁰ Moreover, the limitation is not compelled by the REA, and it does not even fit the best interpretation of that statute.²³¹ The best interpretation of the REA proviso, as well as separation-of-powers limitations, should permit a properly designed sampling procedure when it is needed to enforce the substantive law properly. When this is so, sampling serves procedural goals: it assures

better interpretation of *Wal-Mart* is that the Court compared the skewed outcomes from sampling directly to the parties' substantive law rights. On this interpretation, sampling might well offend the REA or separation-of-powers principles even if it were used in individual suits, again because it systematically skews outcomes away from substantive entitlements.

229. *Id.* at 1046 (holding that the permissibility of sampling in an individual suit is both necessary and sufficient for its use in a class action).

230. *See supra* note 198 and accompanying text. In fact, the Court's argument begs the central question: why sampling does not violate the REA or separation-of-powers principles when it is used in an *individual suit*? Neither *Tyson Foods* nor *Mt. Clemens* addresses this question. I argue in the text that sampling does not violate the REA or offend separation-of-powers principles, and the reason applies equally to class actions and individual suits.

231. The Supreme Court has never articulated a coherent doctrinal framework for applying the REA proviso. Early on, the Court adopted a test that asked whether the Federal Rule of Civil Procedure “really regulates procedure,—the judicial process for enforcing rights and duties recognized by substantive law and for justly administering remedy and redress for disregard or infraction of them.” *Sibbach v. Wilson & Co., Inc.*, 312 U.S. 1, 14 (1941); *see also* *Hanna v. Plumer*, 380 U.S. 460, 464 (1965) (relying on the same test). The Court, however, did not explain how “really” regulating procedure differs from just regulating it. *See Sibbach*, 312 U.S. at 14. In 2010, a plurality of the Court endorsed a version of this test focusing on whether the rule on its face regulates a matter of procedure, regardless of its substantive effects. *Shady Grove Orthopedic Assocs., P.A. v. Allstate Ins. Co.*, 559 U.S. 393, 406–10 (2010) (plurality opinion) (upholding the validity of Rule 23 against a challenge that it enlarged the substantive rights created by state law and embodied in the state claim). And in 1987, the Court made clear that “incidental” substantive effects would not invalidate a Federal Rule of Civil Procedure, especially if the rule was “reasonably necessary to maintain the integrity of [the] system of rules[.]” and also recognized a strong presumption in favor of the validity of a Federal Rule of Civil Procedure based on the careful review that those rules receive during the rulemaking process. *Burlington N.R.R. Co. v. Woods*, 480 U.S. 1, 5, 8 (1987). One might argue that Rule 23 with sampling really regulates procedure because it has a procedural purpose and that its substantive effects are merely incidental. However, one might also argue that sampling converts Rule 23 into more than a joinder device by injecting a substantive dimension and that its substantive effects are more than incidental. It is difficult to resolve this controversy without a clearer understanding of when a Federal Rule of Civil Procedure actually regulates procedure on its face and what constitutes an incidental effect.

that individuals with meritorious claims can obtain a fair and just determination of their claims.²³²

It is true that sampling affects outcomes in predictable and systematic ways, but so do many other procedures. For example, tightening a pleading standard predictably and systematically increases the error risk for plaintiffs with meritorious suits and reduces the error risk for defendants.²³³ And relaxing the pleading standard, as notice pleading does, has predictable and systematic effects in the opposite direction. It follows that outcome effects, even predictable and systematic effects, alone cannot be enough to trigger the REA or separation-of-powers concerns. As long as sampling serves procedural goals—such as distributing the risk of error more fairly and efficiently or reducing process costs—and as long as there is no legislative bar to its use,²³⁴ its substantive effects are merely “incidental” and sampling does not intrude on the legislative sphere.²³⁵

232. See Sergio J. Campos, *Mass Torts and Due Process*, 65 VAND. L. REV. 1059, 1118–21 (2012); see also Robert G. Bone, *The Process of Making Process: Court Rulemaking, Democratic Legitimacy, and Procedural Efficacy*, 87 GEO. L.J. 887, 950–54 (1999) (proposing a test to measure validity based on the rule’s justification); Tobias Barrington Wolff, *Managerial Judging and Substantive Law*, 90 WASH. U. L. REV. 1027, 1046–47 (2013) (arguing that the *Wal-Mart* Court should have considered the question whether sampling was justified as an exercise of federal common law power to provide remedies adequate to further Title VII policies).

233. See Bone, *supra* note 175, at 925–28.

234. If the best interpretation of the federal statute creating the substantive claim clearly required case-specific determinations incompatible with the use of sampling, then sampling would alter substantive rights and violate the REA and separation-of-powers principles.

235. See *Burlington N.R.R. Co.*, 480 U.S. at 5 (holding that “incidental” effects on substantive rights do not trigger the REA proviso). My discussion ignores potential Article III concerns. The *Tyson Foods* Court did not decide whether Article III’s case or controversy requirement prevents a class action from being certified when it includes class members with no legal injury or when those class members cannot be identified and excluded from sharing in the class recovery. See *Tyson Foods, Inc. v. Bouaphakeo*, 136 S. Ct. 1036, 1049–50 (2016); see also *Spokeo, Inc. v. Robins*, 136 S. Ct. 1540, 1547–48 (2016) (holding that the named representative of a class seeking statutory damages must allege concrete injury for Article III standing but stating nothing about whether absent class members must also satisfy standing requirements). This Article III issue is not strictly about sampling; it is about the permissibility of aggregations that include uninjured individuals who would have no Article III standing to bring individual suits. Compare *Neale v. Volvo Cars of N. Am., LLC*, 794 F.3d 353, 364 (3d Cir. 2015) (holding that only the named representative needs Article III standing), with *Denney v. Deutsche Bank AG*, 443 F.3d 253, 264 (2d Cir. 2006) (holding that Article III applies to all class members). In any case, if Article III bars certification of class actions that predictably include uninjured class members, then it also bars sampling whether it is used in a class action or an individual suit.

One might argue that sampling is too substantive for a different reason, namely, that it is likely to be politically controversial and therefore should be vetted through the political process.²³⁶ But this objection goes too far. Taken seriously, it would cripple the court rulemaking process. For example, Rule 23 would violate the REA proviso viewed in this way, as would the Federal Rules dealing with pleading, discovery, and summary judgment, all of which attract intense political controversy these days. The result—that virtually all procedural law in federal court would have to be adopted by Congress—is not only undesirable as a policy matter, but is also plainly inconsistent with the congressional decision in the REA to delegate the bulk of procedural rulemaking to the Supreme Court.²³⁷

3. Sampling Is Too Mechanical

Statistical adjudication substitutes a mechanical and formulaic decision-making protocol for case-specific reasoning. The judge calculates the average (or other statistic) based on the sample results and applies it to all other cases in the aggregation. It is true that the judge employs a reasoning process when she decides whether to sample. Also, the sampled cases are decided in the usual way, by applying reason to the facts in light of the relevant law. But the rest of the aggregation is adjudicated mechanically without applying reason to the facts of each case.

I believe this observation captures something important about adjudication, but we need to be clear about what exactly that is. To unpack the intuition, consider a hypothetical in which a judge flips a coin to decide an issue. Suppose that the issue is committed entirely to the judge's discretion, such as the length of a sentence in a criminal case within the bounds of sentencing guidelines. After hearing extensive discussion, the judge comes to the conclusion that neither side has a more compelling argument. As a result, she proposes to flip a coin. Even if all the parties agree and there are no substantial third party effects, I believe that many, perhaps most, people would

236. See, e.g., Paul D. Carrington, "Substance" and "Procedure" in the Rules Enabling Act, 1989 DUKE L.J. 281, 308 (noting that a rule does not "abridge, enlarge, or modify" substantive rights if it is so general that it does not evoke political controversy).

237. See Stephen B. Burbank, *The Rules Enabling Act of 1934*, 130 U. PA. L. REV. 1015, 1106–12 (1982) (arguing that the principal purpose of the REA proviso was to allocate power between Congress and the Court).

condemn the judge for deciding the issue in this way.²³⁸ When pressed for an explanation, critics are likely to argue that flipping a coin is not a legitimate way for a judge to decide, that judges are supposed to employ reason and not rely on chance.

Notice that the problem is not arbitrariness: the judge in our hypothetical has a sound justification for using a coin flip.²³⁹ From an outcome perspective, the coin flip gives each side an equal chance, which is about as much as parties can expect when a case is in equipoise and there is no burden assignment rule to break ties. Moreover, flipping a coin has the advantage of purging the decision of any possible bias.

The reason that the coin flip is unacceptable, I believe, is not that it produces an arbitrary result, but rather that it displaces the usual case-specific reasoning process at the point of decision.²⁴⁰ One of adjudication's core elements, perhaps its most important, is its commitment to reasoning from general principle in a way that engages the facts of particular cases.²⁴¹ More specifically, judges interpret the law as they apply it by placing existing legal norms alongside the facts of a case. This is a fancy way of describing something very familiar to lawyers: common law reasoning. In the common law system, even clear and seemingly categorical rules are linked to and conditioned on general principles and the facts of

238. See, e.g., *LaPine Tech. Corp. v. Kyocera Corp.*, 130 F.3d 884, 891 (9th Cir. 1997) (Kozinski, J., concurring) (noting that he would refuse to enforce an arbitration clause that “provided that the district judge would review the [arbitration award] by flipping a coin”) *overruled by* *Kyocera Corp. v. Prudential-Bache Trade Servs., Inc.*, 341 F.3d 987 (9th Cir. 2003); *In re Brown*, 662 N.W.2d 733, 734, 736 (Mich. 2003) (accepting the recommendation that a judge be censured for flipping a coin when neither side's argument was more persuasive); Adam M. Samaha, *Randomization in Adjudication*, 51 WM. & MARY L. REV. 1, 28–29, 28 nn.105–06 (2009) (providing several examples of judges being sanctioned for flipping a coin to decide a case).

239. See Lewis A. Kornhauser & Lawrence G. Sager, *Just Lotteries*, 27 SOC. SCI. INFO. 483, 495–505 (1988) (discussing equal entitlement and scarcity conditions for using the lottery as an exclusive or nonexclusive method of allocation and noting that using the lottery under these conditions is supported by reason).

240. See Shay Lavie, Note, *Reverse Sampling: Holding Lotteries to Allocate the Proceeds of Small-Claims Class Actions*, 79 GEO. WASH. L. REV. 1065, 1084–85 (2011) (“The public has every right to expect that a jurist will carefully weigh the matters at issue and . . . render reasoned rulings and decisions.” (quoting N.Y. STATE COMM'N ON JUDICIAL CONDUCT, 1984 ANNUAL REPORT 88)); see also JON ELSTER, *SOLOMONIC JUDGEMENTS: STUDIES IN THE LIMITATIONS OF RATIONALITY* 38 (1989) (emphasizing that “the use of lotteries to resolve decision problems under uncertainty presupposes an unusual willingness to admit the insufficiency of reason”).

241. Robert G. Bone, *Party Rulemaking: Making Procedural Rules Through Party Choice*, 90 TEX. L. REV. 1329, 1387–91 (2012) (defending this point).

previous cases in a way that gives them a more or less flexible quality.²⁴² A judge must decide how strictly to apply an established rule to a particular case—whether to modify the rule or carve out exceptions—by considering predictability, consistency, and other rule-of-law values.²⁴³ The same is true for constitutional law, which develops in a way rather similar to the common law process. And even statutory rules require interpretation to determine whether they apply to the case at hand and what they require on the particular facts.

One way to describe this process is in terms of reflective equilibrium.²⁴⁴ The judge moves back and forth between her best understanding of the law including the principles and policies that support it, and the moral and practical intuitions generated by engaging the facts of the case, all the while adjusting law and intuition until they fit together in a reflective equilibrium.²⁴⁵ My claim then is that this reasoning process is a core, perhaps the core, feature of adjudication. It is essential to the integrity of the institution, and it is what distinguishes adjudication from arbitration, legislation, and other modes of dispute resolution and lawmaking.

This claim is not essentialist; it does not assume a natural law ideal for adjudication. It focuses on civil adjudication as that process is actually practiced in the United States. All it assumes is that the institution has evolved to perform certain functions, such as deciding claims of right, regulating society, and making common law, and that it developed features along the way to help it serve those functions well. Thus, the claim is about *our* system of civil adjudication properly

242. See, e.g., RONALD DWORKIN, *LAW'S EMPIRE* 225–75 (1986) [hereinafter *DWORKIN, LAW'S EMPIRE*]; RONALD DWORKIN, *TAKING RIGHTS SERIOUSLY* 71–80, 110–23, 340–41 (1978); Lon L. Fuller, *The Forms and Limits of Adjudication*, 92 *HARV. L. REV.* 353, 372–81 (1978).

243. Scholars disagree about these matters. Compare *DWORKIN, LAW'S EMPIRE*, *supra* note 242, at 264–66 (noting that judges should sometimes apply rules strictly but always subject to the demands of principled integrity), with FREDERICK SCHAUER, *PLAYING BY THE RULES: A PHILOSOPHICAL EXAMINATION OF RULE-BASED DECISION-MAKING IN LAW AND IN LIFE* 196–206 (1991) (arguing for a strong presumption in favor of strict application).

244. On the method of reflective equilibrium, see generally JOHN RAWLS, *A THEORY OF JUSTICE* 42–45 (rev. ed. 1999).

245. See *DWORKIN, LAW'S EMPIRE*, *supra* note 242, at 225–27 (describing a process of legal reasoning grounded in an effort to form “the best constructive interpretation of the community’s legal practice”); RAWLS, *supra* note 244, at 17–19 (describing the process of seeking a reflective equilibrium).

understood. More precisely, it is about the best normative understanding of that institution.

Indeed, it is quite common to view adjudication as linked to a special type of reasoning process that combines general principles with case-specific facts.²⁴⁶ If this reasoning process is essential to what adjudication does—and I believe it is—then we should worry when a procedure threatens a court’s ability to engage in the process effectively. Deciding a case by flipping a coin falls into this category, as does choosing to have a judge decide a case by reading the entrails of a chicken or empanelling a jury of twelve orangutans.²⁴⁷ These methods of decision would be rejected as illegitimate despite the parties’ agreement, not just because they are bizarre but because they rob the judge of the ability to apply reason to the legally relevant facts.

Although the question of legitimacy is more difficult for sampling, there are some easy cases. The use of sampling to determine facts does not interfere with the reflective equilibrium process in any significant way. *Tyson Foods* is a good example. In that case, sampling was used to determine how long employees took to don and doff their protective equipment. This is as close to a pure fact as one can get. To be sure, facts are inputs into the reasoning process, but innovative methods for finding facts do not trump or distort that process; they just affect the result. Moreover, the relevant FLSA law in *Tyson Foods* was clear and the method of determining liability fairly mechanical—add donning and doffing time to gang-time and subtract K-code time.²⁴⁸ So, there was not much cause to insist on case-specific reflective-equilibrium reasoning.

246. See, e.g., DWORKIN, *LAW’S EMPIRE*, *supra* note 242, at 225–27, 245–47, 254–58; Fuller, *supra* note 242, at 364, 366 (noting that “[a]djudication is . . . a device which gives formal and institutional expression to the influence of reasoned argument in human affairs” and that “the distinguishing characteristic of adjudication lies in the fact that it confers on the affected party a peculiar form of participation . . . that of presenting proofs and reasoned arguments for a decision in his favor”).

247. See *LaPine Tech. Corp. v. Kyocera Corp.*, 130 F.3d 884, 891 (9th Cir. 1997) (Kozinski, J., concurring) (noting that he would refuse to enforce an arbitration clause that “provided that the district judge would review the [arbitration award] by . . . studying the entrails of a dead fowl!”) *overruled by* *Kyocera Corp. v. Prudential-Bache Trade Servs., Inc.*, 341 F.3d 987 (9th Cir. 2003); *United States v. Josefik*, 753 F.2d 585, 588 (7th Cir. 1985) (noting that an agreement to trial by twelve orangutans would be invalid).

248. See *supra* notes 43–46 and accompanying text. Moreover, in a jury trial case, factual determinations are for the jury to decide and we do not expect jurors to engage in a process of reflective equilibrium. To be sure, some jury questions, such as whether the defendant was negligent, involve the application of law to fact, but we assume that the judge first decides the applicable law and then instructs the jury what law to apply. I

Beyond the easy cases, the legitimacy analysis gets murkier. But there are some reasonably clear points. First, it is not necessary that every individual case be decided by a process of reflective equilibrium. If it were, the class action would not be legitimate. Nor would it be legitimate to decide a common question of law and fact separate from the specific facts of individual cases that give it texture and context.

Second, mixed questions of law and fact can be appropriate for sampling. For example, courts are comfortable using approximation methods, including sampling, to estimate damages even though the estimation process can involve the application of law to fact.²⁴⁹ After the *Wal-Mart* decision, for example, some courts continued to use sampling to calculate damages, distinguishing *Wal-Mart* as a case involving sampling for liability.²⁵⁰

These two points go a long way to support the legitimacy of sampling for liability as well as for damages. Sampling preserves the reflective-equilibrium reasoning process for the sampled cases, which is significant because not all cases need be decided in the conventional way. Moreover, the fact that the liability issues involve

ignore the judge-jury distinction in the text and assume that the judge is deciding the issue that is subject to sampling.

249. It is well settled, especially when the plaintiff faces evidentiary or other practical obstacles to proof, that she need not provide a precise measure of damages; it is enough if the plaintiff proposes a method that permits a “‘just and reasonable inference’ . . . even though the result is only an approximation.” *Cristallina S.A. v. Christie, Manson & Woods Int’l, Inc.*, 502 N.Y.S.2d 165, 173 (App. Div. 1986) (quoting *Story Parchment Co. v. Paterson Parchment Paper Co.*, 282 U.S. 555, 563 (1931)); *see Pulaski & Middleman, LLC v. Google, Inc.*, 802 F.3d 979, 989 (9th Cir. 2015) (“In calculating damages, here restitution, California law ‘requires only that some reasonable basis of computation of damages be used, and the damages may be computed even if the result reached is an approximation.’” (quoting *Marsu, B.V. v. Walt Disney Co.*, 185 F.3d 932, 938–39 (9th Cir. 1999)), *cert. denied*, 136 S. Ct. 2410 (2016) (mem.); *Sw. Stainless, LP v. Sappington*, 582 F.3d 1176, 1186 (10th Cir. 2009) (noting that “[w]hile the amount of lost profits ‘may not be based on mere speculation, conjecture and surmise alone, the mere uncertainty as to the exact amount of damages will not preclude the right of recovery. It is sufficient if the evidence shows the extent of damage by just and reasonable inference.’” (quoting *Florafax Int’l, Inc. v. GTE Mkt. Res., Inc.*, 933 P.2d 282, 296 (Okla. 1997))). And as *Tyson Foods* confirms, statistical sampling can be part of any such method. *Tyson Foods, Inc. v. Bouaphakeo*, 136 S. Ct. 1036, 1047 (2016); *see also* 15 U.S.C. § 15d (2012) (allowing damages to “be proved and assessed in the aggregate by statistical or sampling methods” in a *parens patriae* price-fixing case once the defendant has been found liable for a price-fixing agreement).

250. *See, e.g., In re Urethane Antitrust Litig.*, 768 F.3d 1245, 1256–57 (10th Cir. 2014) (distinguishing *Wal-Mart* on the ground that the sampling in that case would have decided liability issues and noting that “*Wal-Mart* does not prohibit certification based on the use of extrapolation to calculate damages”).

the application of law to fact is not in itself disqualifying. Judges decide common questions independent of case-specific facts today, and sampling is used to determine damages even when the inquiry involves law applied to fact.

There is a third important point. Legitimacy cannot be just a function of the decision-making method. Outcome quality matters, too. In the extreme coin flip, chicken entrails, and orangutan examples above, the adverse effect on adjudicative reasoning was clear, direct, and serious, and there was no particular reason to use the procedure other than to satisfy party preferences. By contrast, sampling preserves the reasoning process in the sampled cases and furthers the purposes of adjudication by improving efficiency and making outcomes more fair and just for all parties.

These considerations also impose significant limits on the use of sampling. Judges cannot just aggregate any set of related cases and decide them all by sampling a few. Not only must the aggregation be sufficiently homogeneous, but there must be a sound justification for sampling from it, a justification that meets rights-based as well as utilitarian constraints. And the sampling procedure must not sacrifice the usual process of adjudicative reasoning for too many cases.²⁵¹

Finally, while it is not essential for legitimacy, it might be wise, insofar as possible, to have sampled cases decided by different courts. Doing so has the advantage of harvesting a range of views on the issues being decided.²⁵² In addition, the decisions in each sampled case should have precedential force in future cases to the extent decisions ordinarily do. These measures would ensure that the sampled cases are treated not just as samples for extrapolation purposes, but also as ordinary cases contributing to the adjudicative process in the usual way.

251. For example, suppose that judges in some state choose to routinely decide negligence cases by sampling. These judges wait until enough automobile accident, medical malpractice, or other negligence cases are filed and then sample from all the filed cases, giving the rest the average of the sample verdicts. There are many different reasons to object to this procedure, including concerns about population heterogeneity, the magnitude of social benefits, the treatment of rights, and legitimacy. And from a legitimacy perspective, it is significant that deciding cases in this way jettisons case-specific reasoning in a huge swath of cases without any special justification for doing so.

252. The district court conducting the sampling would have to have authority to transfer sampled cases to other courts, which might require legislation similar to the Multidistrict Litigation Act. *See* 28 U.S.C. § 1407 (2012).

CONCLUSION

There is reason to celebrate the *Tyson Foods* decision and to read it broadly. Statistical adjudication through sampling can be a very useful procedural tool. Like summary judgment, case management, and other procedures, it helps assure that adjudication accomplishes its goals in the best possible way. Sampling must be used properly, and this Article develops a set of factors to guide decisions about whether and how to use it. These factors take account of what the *Tyson Foods* Court says and what it does in light of the broader purposes of civil adjudication.

Still, many are likely to balk at the idea that courts can decide cases by extrapolating from a sample. Even the *Tyson Foods* Court refuses to acknowledge that the case is one of statistical adjudication, choosing instead to pretend that it is merely a case of statistical evidence. I argued that these reservations have to do with nagging doubts about the legitimacy of the procedure. I then examined three legitimacy objections: that statistical adjudication is too strange, that it is too substantive, and that it is too mechanical. None of these objections is decisive. When used properly, statistical adjudication through sampling neither transgresses the legitimate bounds of civil adjudication, nor violates the substantive or procedural rights of parties. It is past time that we embrace statistical adjudication without skepticism or hostility. *Tyson Foods* opens the door. It is up to us to seize the opportunity.

