

CARDOZO LAW REVIEW
de•novo

SAMPLING AND RELIABILITY IN CLASS ACTION
LITIGATION

Hillel J Bavli[†]

TABLE OF CONTENTS

INTRODUCTION	207
I. RELIABILITY IN THE LAW	209
II. CLAIM AND JUDGMENT VARIABILITY	210
III. REDUCING JUDGMENT VARIABILITY WITH SAMPLING.....	211
IV. CONDITIONS UNDER WHICH SAMPLING CAN IMPROVE THE RELIABILITY OF LEGAL OUTCOMES.....	212
V. IMPLICATIONS AND CONCLUSIONS	214

INTRODUCTION

Courts continue to struggle with the limits of statistical sampling in resolving claims arising from “the mass repetitive wrong.”¹ In *Cimino v. Raymark Indus., Inc.*, a 1990 asbestos class action, U.S. District Court Judge Robert Parker divided a class of 2,298 claimants into five

[†] Harvard University. The views expressed in this Article are those of the author, and not those of any organization with which he is affiliated. The author thanks John Kenneth Felner for his helpful comments and edits.

¹ *Cimino v. Raymark Indus., Inc.*, 751 F. Supp. 649, 652 (E.D. Tex. 1990) (citing HERBERT B. NEWBERG, *NEWBERG ON CLASS ACTIONS* § 17.06, at 373 (2d ed. 1985)).

subclasses based on claimed injuries, and selected a random sample from each subclass to form a representative sample group of 160 claimants.² He submitted the sample group's claims to a jury for individual determinations and then extrapolated outcomes for the non-sampled claims by applying the average sample group determinations to each subclass, respectively.³ The Fifth Circuit ultimately rejected Judge Parker's sampling procedure on constitutional grounds.⁴

In recent years, courts and commentators have criticized the use of sampling to prove classwide liability and damages.⁵ It is widely believed that sampling serves goals of efficiency, but only at the cost of reliability—and that the “benefits of efficiency can never be purchased at the cost of fairness.”⁶

In a 1992 Stanford Law Review article, Professors Michael Saks and Peter Blanck argued that sampling, when performed correctly, not only satisfies the standards of reliability achievable through individual litigation, but that sampling can also *increase* the reliability of legal outcomes.⁷

However, courts have generally been unwilling to accept the reliability of sampling procedures used to prove classwide liability and damages; and courts frequently reject sampling on constitutional and procedural grounds.⁸ But, as Saks and Blanck have asserted, “a major—perhaps *the* major—due process concern in an aggregated trial is the validity of the outcome.”⁹ Moreover, “[t]he main argument against trial by aggregation and sampling asserts that such trials cannot give the parties as accurate a result as they would obtain through traditional bilateral trials.”¹⁰

² *Id.* at 652–54; see also Hillel J. Bavli, *Aggregating for Accuracy: A Closer Look at Sampling and Accuracy in Class Action Litigation*, 14 LAW, PROBABILITY & RISK 67, 68 (2015) [hereinafter *Aggregating for Accuracy*].

³ *Aggregating for Accuracy*, *supra* note 2, at 68.

⁴ See *Cimino v. Raymark Indus., Inc.*, 151 F.3d 297 (5th Cir. 1998).

⁵ See, e.g., *Wal-Mart Stores, Inc. v. Dukes*, 131 S. Ct. 2541, 2561 (2011).

⁶ *Malcolm v. Nat'l Gypsum Co.*, 995 F.2d 346, 350 (2d Cir. 1993); see also *In re Chevron U.S.A., Inc.*, 109 F.3d 1016, 1023 (5th Cir. 1997) (Jones, J., concurring).

⁷ 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, 815–19 (1992).

⁸ See *Wal-Mart Stores, Inc.*, 131 S. Ct. at 2561; see also *Ortiz v. Fibreboard Corp.*, 527 U.S. 815, 845 (1999) (citing *Amchem Prod., Inc. v. Windsor*, 521 U.S. 591, 613 (1997)).

⁹ Saks & Blanck, *supra* note 7, at 833 (emphasis in original); see also *Joint Anti-Fascist Refugee Comm. v. McGrath*, 341 U.S. 123, 171–72 (1951) (Frankfurter, J., concurring); Martin H. Redish & Lawrence C. Marshall, *Adjudicatory Independence and the Values of Procedural Due Process*, 95 YALE L.J. 455, 476–77 (1986).

¹⁰ Saks & Blanck, *supra* note 7, at 833; see also *In re Chevron U.S.A., Inc.*, 109 F.3d at 1020 (“[O]ur procedural due process concerns focus on the fact that the procedure embodied in the district court’s trial plan is devoid of safeguards designed to ensure that the claims against Chevron of the non-represented plaintiffs as they relate to liability or causation are determined in a proceeding that is reasonably calculated to reflect the results that would be obtained if those claims were actually tried.”).

The question remains: why have courts and commentators largely discounted the argument that sampling offers not only increased efficiency, but reliability as well? In a 2015 paper, I explain that a primary source of skepticism may be the generality with which the argument has been made.¹¹ Although Saks and Blanck, and others, have suggested that sampling may increase the reliability of legal outcomes, the literature has not adequately developed the argument or produced a framework to analyze the effect of sampling on reliability.¹² Indeed, sampling does not inevitably increase the reliability of legal outcomes; rather, its effect on reliability depends on the particular features of the claims.¹³ Therefore, in *Aggregating for Accuracy*, I develop a formal framework for examining conditions under which sampling can increase the reliability of legal outcomes.¹⁴

In this Article, I explain my conclusions in *Aggregating for Accuracy* in non-mathematical terms, and underscore certain implications with respect to class action litigation and considerations in light of the U.S. Supreme Court's recent decision in *Tyson Foods, Inc. v. Bouaphakeo*.¹⁵ I begin by describing the building blocks of my analysis—the concepts of reliability and accuracy in the law.

I. RELIABILITY IN THE LAW

Assume that for every legal claim there is a “correct” outcome that can be determined by applying the “true” state of the law to the complete facts surrounding the claim.¹⁶ But, in the real world, complete knowledge regarding the law and the facts surrounding a claim is unavailable. Consequently, a legal outcome (resulting from a trial or other adversarial proceeding) serves as an *estimate* of the “correct” outcome for the claim. Therefore, for every legal claim, there is an error term that (although generally unknown) reflects the disparity between the observed outcome and the “correct” outcome. *Accuracy*, then, is defined in terms of the proximity of the observed outcome to the “correct” outcome.¹⁷

¹¹ *Aggregating for Accuracy*, *supra* note 2, at 69.

¹² *Id.*

¹³ *Id.*

¹⁴ *See generally id.*

¹⁵ 136 S. Ct. 1036 (2016).

¹⁶ Alternatively, it can be assumed that there is a *distribution* of “correct” outcomes associated with every legal claim. *See generally Aggregating for Accuracy*, *supra* note 2, at 69.

¹⁷ *See generally id.* at 74; Jonathan J. Koehler & Daniel Shavero, *Veridical Verdicts: Increasing Verdict Accuracy Through the Use of Overtly Probabilistic Evidence and Methods*, 75 CORNELL L. REV. 247 (1990) (discussing “accuracy”). Note that *accuracy* can be defined more formally using the concepts of bias and variance. However, an explicit discussion of these concepts is beyond the scope of this Article.

One convenient and arguably sensible conceptualization of the “correct” outcome associated with a legal claim is the award that would result from computing the mean of infinitely many adjudications of the claim under various conditions (e.g., various jury compositions, judges, attorneys, presentations of evidence, etc.).¹⁸

Thus, I define the *reliability* of a legal procedure as the accuracy of the legal outcome that can be expected by following the procedure. For example, if it is expected that a certain legal procedure will produce a highly accurate outcome—an observed outcome in close proximity to the “correct” outcome—the procedure is considered reliable.¹⁹

An in-depth discussion of the role of accuracy in the law is beyond the scope of this Article. I assume that accuracy is a fundamental goal of the law, whether such goal is grounded in accuracy itself or a further aim, such as deterrence or fairness.²⁰

II. CLAIM AND JUDGMENT VARIABILITY

Critics of sampling have focused on the error that results from applying a point estimate—e.g., the average of the sample group outcomes—to a class of heterogeneous legal claims.²¹ Heterogeneity can be described in terms of *claim variability*—i.e., the variability of the facts of the claims, or, more precisely, the variability of the “correct” outcomes associated with individual claims.²²

But critics of sampling have ignored a second type of error: error resulting from *judgment variability*.²³ Judgment variability represents the randomness associated with a claim’s outcome. It arises from the variability in conditions under which outcomes are determined, including the composition of the jury, the judge presiding over the case, the presentation of evidence, the attorneys involved in the case, etc.²⁴ For example, if a single claim is tried ten times independently (each trial

¹⁸ See Saks & Blanck, *supra* note 7, at 833–34; *Aggregating for Accuracy*, *supra* note 2, at 74. Other measures of central tendency, such as the median, are also possible. For simplicity, throughout this Article we use solely the mean.

¹⁹ See *Aggregating for Accuracy*, *supra* note 2, at 74–78 for a detailed discussion of accuracy and variability in the law.

²⁰ See *id.* at 75; Saks & Blanck, *supra* note 7, at 829 (citing Redish & Marshall, *supra* note 9, at 476–77).

²¹ *Aggregating for Accuracy*, *supra* note 2, at 75–76. See generally Saby Ghoshray, *Hijacked by Statistics, Rescued by Wal-Mart v. Dukes: Probing Commonality and Due Process Concerns in Modern Class Action Litigation*, 44 LOY. U. CHI. L.J. 467 (2012). If there are subclasses, then a court is likely to assign the average sample group outcome within each subclass to the non-sampled claims in each subclass, respectively. See, e.g., *Cimino v. Raymark Indus., Inc.*, 751 F. Supp. 649, 651–54 (E.D. Tex. 1990).

²² *Aggregating for Accuracy*, *supra* note 2, at 76.

²³ *Id.* at 76–78.

²⁴ *Id.* at 76–77.

with a new selection of trier of fact, attorneys, etc.), it is likely that there would be ten distinct verdicts. But, if there is a *single correct outcome* associated with the claim, then judgment variability reflects error—disparities between the observed outcomes and the “correct” outcome.²⁵

For simplicity, I assume that, *on average*, a case will result in the “correct” outcome—in statistical terms, the outcome is *unbiased*. Judgment variability represents the degree to which an observed outcome varies around the “correct” outcome.²⁶ The reader may imagine a bell curve centered at the “correct” outcome, where the judgment variability determines the width of the curve. On average, the outcome will be the “correct” outcome; but the outcome is variable around the “correct” outcome, and such randomness, represented by judgment variability, reflects error.²⁷

In determining appropriate standards in the class action context, courts and commentators have focused on error resulting from claim variability, but have generally ignored error resulting from judgment variability.²⁸ Indeed, even the purported ideal of individual litigation produces outcomes that are subject to judgment variability.

III. REDUCING JUDGMENT VARIABILITY WITH SAMPLING

Consider a costly procedure through which the outcome of a claim is determined by averaging the verdicts resulting from ten independent “replications” of a trial, or “repeated adjudications” (involving, for example, different judges, juries, attorneys, presentations of evidence, etc.). Assuming the outcome is relatively unbiased, it is easy to show that following this procedure results in an accurate outcome—an outcome that is close to the “correct” outcome. Similarly, this procedure will produce an accurate outcome for each claim of each member of a putative class (or subclass). Replication thus increases the reliability of legal outcomes by reducing the error caused by judgment variability.

On the other hand, the costs of such a procedure are enormous, and likely not justified by the benefits of the procedure for a single claim (or for each putative class member’s claim that requires application of the procedure for the individual claim).

Importantly, however, in a class of homogeneous claims, it is not necessary to follow the procedure for each class member’s claim to

²⁵ See Saks & Blanck, *supra* note 7; Edward K. Cheng, *When 10 Trials Are Better than 1000: An Evidentiary Perspective on Trial Sampling*, 160 U. PA. L. REV. 955, 958–60 (2012); *Aggregating for Accuracy*, *supra* note 2, at 75–78.

²⁶ See generally *Aggregating for Accuracy*, *supra* note 2, at 75–78.

²⁷ See *id.* at 75–79. Consider the magnitude of error that results from judgment variability when classwide damages are determined by adjudicating a single representative claim.

²⁸ *Id.* at 76–78.

realize the benefits of the procedure: following the replication procedure for a single claim and extrapolating the result to all claims of the homogeneous class (or subclass) will yield the reliability benefits as though the procedure was followed for each claim.

Class actions provide opportunities to realize the reliability benefits of replication without incurring the prohibitive costs that such procedures entail.²⁹ In particular, courts can use sampling to improve legal outcomes by reducing error resulting from judgment variability. Moreover, sampling may offer a degree of reliability that cannot be obtained even through the purported ideal of individual adjudication. Further, although homogeneity is helpful, it is not necessary. As explained below, the benefits of sampling in a heterogeneous class depend on whether, and to what extent, the error resulting from judgment variability dominates the error resulting from claim variability.³⁰

IV. CONDITIONS UNDER WHICH SAMPLING CAN IMPROVE THE RELIABILITY OF LEGAL OUTCOMES

Above, I describe how repeated adjudication can increase the reliability of legal outcomes. Assuming a certain degree of homogeneity, if a court were unconstrained by cost or law in its pursuit of reliability, it might “sample” all the claims of a class for individual adjudication and then replace *all* individual outcomes with a single aggregated outcome. In fact, a court could further reduce the error resulting from judgment variability by adjudicating each claim twice, or more times for that matter. But, in addition to the high costs of litigation, a court’s ability to achieve extremely reliable outcomes is constrained by law.

In *Aggregating for Accuracy*, I argue that a court may not replace an individualized outcome with an aggregated one.³¹ I highlight a fundamental distinction, based on constitutional law and rules of civil procedure, between a court’s authority to extrapolate a representative claim’s determination to *unadjudicated* claims and its authority to *replace an individually adjudicated outcome* with an aggregated one.³² A detailed discussion of this issue is beyond the scope of this Article. Assume, therefore, that a court must choose between sampling a claim for individual adjudication and preserving the claim for assignment of

²⁹ *Id.* at 77.

³⁰ *Id.* at 78; *see* Saks & Blanck, *supra* note 7, at 833–37.

³¹ *Aggregating for Accuracy*, *supra* note 2, at 78–80.

³² *Id.*

an aggregated outcome extrapolated from the sampled claims.³³

This constraint can be modeled using a type of exploration-exploitation tradeoff I call *reductive sampling*, in which sampling a unit—here, a legal claim—means reducing (or, as here, eliminating) its eligibility for later extrapolation with respect to that unit.³⁴

In *Aggregating for Accuracy*, I show that, using the reductive sampling framework and standard statistical tools for minimizing error, in the context of a class of homogeneous (or relatively homogeneous) claims, accuracy is maximized, not by adjudicating each of the claims individually (as is often viewed as ideal), but rather by determining individual outcomes for a small random sample of claims, and assigning the average of the sample group outcomes to the remaining, non-sampled, claims.³⁵ In particular, I show that, for a class of N homogeneous claims, a court maximizes accuracy by sampling \sqrt{N} claims, rather than all N claims, for individual adjudication.³⁶ For example, if the class contains 5,000 homogeneous claims, a court can maximize accuracy by randomly sampling about 70 claims for individual adjudication, assigning the 70 individual outcomes to the sampled claims, and assigning the arithmetic mean of the 70 sample outcomes to the remaining 4,930 claims.

\sqrt{N} (or about 70 in the example above) is the number that balances, on the one hand, a court's interest in obtaining information regarding the "correct" outcome of the homogeneous claims (which increases with sample size), and, on the other hand, its interest in preserving claims to which to assign the *accurate* aggregated outcome.³⁷ The 70 claims in the sample group receive individual adjudications that are subject to significant error caused by judgment variability, whereas the remaining 4,930 claims in the extrapolation group are assigned aggregated outcomes that have been "refined" by repeated adjudication—i.e., outcomes whose judgment variability has been reduced significantly by averaging over approximately 70 repeated adjudications.³⁸

Now, to examine sampling in the context of *heterogeneous* claims, consider an additional factor: although the sampling procedure described above reduces error caused by judgment variability, assigning a single aggregated outcome (e.g., the average of the sample group adjudications) as the estimate of the "correct" outcomes associated with a group of *heterogeneous* claims—claims that actually involve

³³ See *id.* at 80–81.

³⁴ *Id.* See generally Herbert Robbins, *Some Aspects of the Sequential Design of Experiments*, 58 BULL. AM. MATHEMATICAL SOC'Y 527 (1952); J. C. Gittins, *Bandit Processes and Dynamic Allocation Indices*, 41 J. ROYAL STAT. SOC'Y: SERIES B (METHODOLOGICAL) 148 (1979).

³⁵ See *Aggregating for Accuracy*, *supra* note 2, at 81–82.

³⁶ *Id.*

³⁷ See *id.* at 80–81.

³⁸ *Id.* at 80–82.

numerous *distinct* “correct” outcomes—introduces error reflecting the disparities between the estimate and each of the “correct” outcomes.³⁹ Thus, as the heterogeneity of the class increases, the value of assigning a single aggregated outcome as the estimate of the “correct” outcomes decreases.⁴⁰ Further, at some degree of heterogeneity, the benefits of sampling, with respect to judgment variability, are outweighed by the detriments of sampling, with respect to claim variability.⁴¹

In *Aggregating for Accuracy*, I show that if claim variability is zero, then the optimal sample size is \sqrt{N} , the homogeneous optimum; if claim variability is greater than judgment variability (e.g., in terms of damages awarded), the optimal sample size is N , which is equivalent to individual adjudications; and finally, if judgment variability is greater than claim variability, then the optimal sample size is between \sqrt{N} and N , and can be determined by a particular formula (derived in *Aggregating for Accuracy*) involving the number of claims in the class, claim variability, and judgment variability.⁴²

V. IMPLICATIONS AND CONCLUSIONS

The conclusions explained above, and derived in *Aggregating for Accuracy*, have important implications for a court’s treatment of statistical sampling in class action litigation. Perhaps most importantly, a court should not assume, as many courts have, that sampling reduces reliability. The discussion above makes clear that sampling may enhance reliability as well as efficiency.

For example, these conclusions have important implications in the context of class actions brought under Rule 23(b)(3) of the Federal Rules of Civil Procedure. In particular, class representatives have proposed sampling procedures for purposes of fulfilling the Rule’s requirement that “the questions of law or fact common to class members predominate over any questions affecting only individual members.”⁴³ In addressing this requirement, the class representatives have offered sampling-based methodologies at the class certification stage in attempt to demonstrate that liability and damages can be determined on a classwide basis.⁴⁴ Courts regularly reject such attempts, however, on grounds—implicitly or explicitly—of reliability. But, for the reasons

³⁹ *Id.* at 82–83.

⁴⁰ *Id.*

⁴¹ *See id.* at 82–83. *See generally* Saks & Blanck, *supra* note 7.

⁴² *See Aggregating for Accuracy*, *supra* note 2, at 83.

⁴³ *See Comcast Corp. v. Behrend*, 133 S. Ct. 1426, 1432 (2013) (quoting FED. R. CIV. P. 23(b)(3)).

⁴⁴ *See Comcast Corp.*, 133 S. Ct. at 1433–34; *Tyson Foods, Inc. v. Bouaphakeo*, 136 S. Ct. 1036, 1043–44 (2016).

discussed above, courts are generally not justified in assuming that sampling diminishes reliability.

The Supreme Court, in its recent decision in *Tyson Foods, Inc. v. Bouaphakeo*, refused to adopt “a broad rule against the use in class actions of . . . representative evidence”—specifically, evidence based on “a representative or statistical sample” offered to establish classwide liability.⁴⁵ Instead, it held that “[w]hether and when statistical evidence can be used to establish classwide liability will depend on the purpose for which the evidence is being introduced and on ‘the elements of the underlying cause of action.’”⁴⁶

Tyson Foods involved claims by Tyson employees, certified by the district court as a Rule 23(b)(3) class, alleging that their employer, Tyson, failed to pay compensable overtime wages under the Fair Labor Standards Act (FLSA) and Iowa law for time spent “donning and doffing protective gear.”⁴⁷ The district court submitted the issues of liability and damages to the jury, which “returned a special verdict finding that time spent in donning and doffing protective gear . . . was compensable work,” and “awarded the class about \$2.9 million in unpaid wages.”⁴⁸ Tyson appealed, arguing, *inter alia*:

[T]he class should not have been certified because the primary method of proving injury assumed each employee spent the same time donning and doffing protective gear, even though differences in the composition of that gear may have meant that, in fact, employees took different amounts of time to don and doff.⁴⁹

The Eighth Circuit Court of Appeals affirmed the judgment and the award.⁵⁰

In ruling that the district court did not err in certifying the class, the Supreme Court held:

In many cases, a representative sample is “the only practicable means to collect and present relevant data” establishing a defendant’s liability. *Manual of Complex Litigation* §11.493, p. 102 (4th ed. 2004). In a case where representative evidence is relevant in proving a plaintiff’s individual claim, that evidence cannot be deemed improper merely because the claim is brought on behalf of a class. To so hold would ignore the Rules Enabling Act’s pellucid instruction that use of the class device cannot “abridge . . . any

⁴⁵ *Tyson Foods, Inc.*, 136 S. Ct. at 1046.

⁴⁶ *Id.* (quoting *Erica P. John Fund, Inc. v. Halliburton Co.*, 563 U.S. 804, 809 (2011)).

⁴⁷ *Tyson Foods, Inc.*, 136 S. Ct. at 1042.

⁴⁸ *Id.* at 1044. Note that a class expert recommended an award of approximately \$6.7 million. *Id.* at 1052.

⁴⁹ *Id.* at 1041.

⁵⁰ *Id.* at 1044; *Bouaphakeo v. Tyson Foods, Inc.*, 765 F.3d 791 (8th Cir. 2014).

substantive right.” 28 U.S.C. § 2072(b).⁵¹

The Court continued: “One way for respondents to show, then, that the sample relied upon here is a permissible method of proving classwide liability is by showing that each class member could have relied on that sample to establish liability if he or she had brought an individual action.”⁵² It held that “[i]f the sample could have sustained a reasonable jury finding as to hours worked in each employee’s individual action, that sample is a permissible means of establishing the employees’ hours worked in a class action.”⁵³

The Court clarified its ruling in *Wal-Mart Stores, Inc. v. Dukes*,⁵⁴ explaining that “*Wal-Mart* does not stand for the broad proposition that a representative sample is an impermissible means of establishing classwide liability.”⁵⁵ *Wal-Mart* involved a class of approximately 1.5 million current and former female employees alleging gender discrimination under Title VII.⁵⁶ The Court explained that “[t]he plaintiffs in *Wal-Mart* did not provide significant proof of a common policy of discrimination to which each employee was subject.”⁵⁷ Ultimately, the Supreme Court rejected plaintiffs’ proposed methodology—proposed “as a means of overcoming th[e] absence of a common policy”⁵⁸—by which a sample of class members would be selected for individual determinations of liability and backpay, and the aggregated damages award would be derived by extrapolating the “number of (presumptively) valid claims” from the percentage of the *sampled* claims determined to be valid, and then multiplying this number by “the average backpay award in the sample set.”⁵⁹

The Court explained that its holding in *Tyson Foods* is “in accord with” its decision in *Wal-Mart*: “Since the Court held that the employees were not similarly situated, none of them could have prevailed in an individual suit by relying on depositions detailing the ways in which other employees were discriminated against by their particular store managers.”⁶⁰ The Court explained that “[b]y extension,

⁵¹ *Tyson Foods, Inc.*, 136 S. Ct. at 1046.

⁵² *Id.*

⁵³ *Id.* at 1046–47. The Court explained:

If the employees had proceeded with 3,344 individual lawsuits, each employee likely would have had to introduce [the class expert’s] study to prove the hours he or she worked. Rather than absolving the employees from proving individual injury, the representative evidence here was a permissible means of making that very showing.

Id. at 1047.

⁵⁴ 564 U.S. 338 (2011).

⁵⁵ *Tyson Foods, Inc.*, 136 S. Ct. at 1048.

⁵⁶ *Id.*

⁵⁷ *Id.*

⁵⁸ *Id.*

⁵⁹ *Id.* (quoting *Wal-Mart Stores, Inc. v. Dukes*, 564 U.S. 338, 367 (2011)).

⁶⁰ *Tyson Foods, Inc.*, 136 S. Ct. at 1048.

if the employees had brought 1½ million individual suits, there would be little or no role for representative evidence,” and that “[p]ermitting 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.”⁶¹ In *Tyson Foods*, the Court held:

[T]he study here could have been sufficient to sustain a jury finding as to hours worked if it were introduced in each employee’s individual action. While the experiences of the employees in *Wal-Mart* bore little relationship to one another, in this case each employee worked in the same facility, did similar work, and was paid under the same policy.⁶²

Thus, although the Supreme Court explicitly refused to adopt “broad and categorical rules governing the use of representative and statistical evidence in class actions,” the *Tyson Foods* decision approves the use of representative evidence to establish classwide liability.⁶³

Distinguish two types of sampling methodologies that putative class representatives have attempted to use for establishing classwide liability and damages: 1) the use of “representative evidence” offered to the trier of fact as probative of classwide liability⁶⁴ or damages, and 2) the use of representative adjudications to extrapolate outcomes for non-sampled (i.e., non-adjudicated) claims.⁶⁵ Although the Supreme Court did not distinguish between these two types of sampling methodologies, it is likely that *Tyson Foods* expands the ability of class representatives to use *representative evidence* to establish classwide liability in particular.

⁶¹ *Id.*

⁶² *Id.* at 1048. The Supreme Court held that, although the “question whether uninjured class members may recover is one of great importance,” it is not “a question yet fairly presented by this case, because the damages award has not yet been disbursed, nor does the record indicate how it will be disbursed.” *Id.* at 1050. The Court remanded the case for further proceedings consistent with its opinion. *Id.* Significantly, Chief Justice Roberts expressed concern, in a concurring opinion joined in part by Justice Alito, that, since the district court may be unable to “fashion a method for awarding damages only to those class members who suffered an actual injury,” *id.* at 1050, “it remains to be seen whether the jury verdict can stand.” *Id.* at 1053. Additionally, Justice Thomas, in a dissenting opinion joined by Justice Alito, asserted that “[t]he District Court erred at the class certification stage by holding that the plaintiffs satisfied Rule 23’s predominance requirement.” *Id.* at 1054. According to the dissenting opinion, the issue of “whether each employee worked over 40 hours without receiving full overtime pay” was “clearly individualized,” and, with respect to the “critical issue” of whether the “individualized nature of employees’ donning and doffing times defeated predominance,” *id.*, the district court erred by certifying the class without giving “proper consideration to the significance of variable donning and doffing times.” *Id.* at 1055.

⁶³ *Id.* at 1049.

⁶⁴ *See id.* at 1043.

⁶⁵ *See Aggregating for Accuracy*, *supra* note 2, at 70–72 (citing cases and literature); *see, e.g., Cimino v. Raymark Indus., Inc.*, 751 F. Supp. 649, 652–54 (E.D. Tex. 1990).

Regarding the use of representative evidence, the Court held:

A representative or statistical sample, like all evidence, is a means to establish or defend against liability. Its permissibility turns not on the form a proceeding takes—be it a class or individual action—but on the degree to which the evidence is *reliable* in proving or disproving the elements of the relevant cause of action.⁶⁶

The Court's decision in *Tyson Foods* approves the use of statistical sampling to establish classwide liability; but the role of *reliability* in determining whether a court permits statistical sampling to prove classwide liability remains central to the analysis.

The conclusions explained above suggest that, while a court should examine the reliability of statistical sampling for, among other things, methodological flaws and issues related to the cohesiveness of the class, it should not discount the reliability of statistical sampling—and representative evidence in particular—because of the sampling itself. Indeed, as explained above, sampling may improve reliability as well as efficiency.

Additionally, it is important to realize that, although the discussion above relates particularly to the reliability benefits of representative adjudications, the conclusions generally apply to representative *evidence* as well. As explained, repeated adjudication may improve reliability by enabling a court, in essence, to incorporate additional information regarding a class of claims—e.g., by averaging over multiple adjudications rather than relying on a single adjudication—and thus minimize error caused by judgment variability. Repeated adjudication in a heterogeneous class similarly confers reliability benefits, as long as the error-reducing benefits of “information sharing” with respect to judgment variability outweigh the error-inducing costs with respect to claim variability.

The use of representative evidence similarly improves reliability. Although the method of “information sharing” using representative evidence is different—involving, for example, the additional step of providing the information to the trier of fact, rather than incorporating it in the outcome directly (e.g., by averaging over repeated adjudications)—insofar as the trier of fact incorporates the representative evidence in its determination of the outcome, the reliability benefits of repeated adjudication apply similarly to the use of representative evidence.⁶⁷

⁶⁶ *Tyson Foods, Inc.*, 136 S. Ct. at 1046 (citing FED. R. EVID. 401, 403, 702) (emphasis added).

⁶⁷ In particular, in *Aggregating for Accuracy*, an award in a heterogeneous class is modeled hierarchically: the “correct” awards in the class are distributed around some global mean, whereas each actual award is “drawn” from a distribution around each claim's “correct” award. See generally *Aggregating for Accuracy*, *supra* note 2, at 82–83. The former distribution represents

Tyson Foods is likely to have a significant impact on class action litigation. Putative class representatives will be encouraged to use representative evidence to establish classwide liability, and perhaps damages as well. Using arguments establishing, for example, that “each class member could have relied on th[e] sample to establish liability if he or she had brought an individual action,” putative class representatives will improve their ability to establish predominance under Rule 23(b)(3).⁶⁸

In light of *Tyson Foods*, a court considering Rule 23(b)(3) certification is likely to focus more heavily and more explicitly on the reliability of representative evidence and statistical sampling generally. Although a statistical sample should be carefully scrutinized to detect, among other things, methodological deficiencies and issues regarding the cohesiveness of the class, sampling often improves the reliability of legal outcomes.

claim variability, whereas the latter distribution represents judgment variability. In a homogeneous class, replication offers accuracy benefits by providing additional information regarding the “correct” award associated with the replicated claim, which otherwise would be obscured by judgment variability. In a heterogeneous class, sampling offers accuracy benefits, with respect to a certain claim, not by providing information regarding that claim’s “correct” award directly, but by providing information regarding the global mean around which all of the “correct” awards are distributed, and thereby regarding the “correct” award for the subject claim indirectly. Similarly, representative evidence, such as the type in dispute in *Tyson Foods* (where, for example, the sample reflects variability of measured donning and doffing times rather than judgment variability), offers accuracy benefits, with respect to a certain claim, by providing information regarding the global mean around which the “correct” awards are distributed, and thereby regarding the “correct” award for that claim in particular. Another way of understanding this is through “comparable-case guidance” (CCG) methods, whereby a court uses information regarding awards in prior comparable cases as guidance for a fact-finder’s determination of damages. See Hillel J. Bavli, *The Logic of Comparable-Case Guidance in the Determination of Awards for Pain and Suffering and Punitive Damages*, U. CIN. L. REV. (forthcoming 2017) [hereinafter *The Logic of CCG*]. *The Logic of CCG* examines the statistical mechanism by which CCG affects awards, and the conditions under which such evidence will improve accuracy. In particular, the paper explains that, under certain mild behavioral assumptions, the risk that such evidence would reduce accuracy—that error resulting from claim variability and bias would outweigh the accuracy benefits of reducing judgment variability—is minimal. See *id.* Like CCG, representative evidence provides information regarding the distribution of “correct” awards for comparable claims, including the global mean, and, in turn, about the “correct” award for the subject claim.

⁶⁸ *Tyson Foods, Inc.*, 136 S. Ct. at 1046.