

Jury Poker: A Statistical Analysis of the Fair Cross-Section Requirement

Richard M. Ré*

In describing the Sixth Amendment's requirement that criminal juries must be drawn from a fair cross-section of the community,¹ the Supreme Court has discussed jury selection as though it were a kind of poker. In this analogy, jurors of "distinctive" backgrounds—typically women, African Americans, or Hispanics—are likened to playing cards of a particular suit.² Just as the rules of any serious poker game demand a complete and fairly shuffled deck, the fair cross-section requirement, according to the Supreme Court, "deprives the State of the ability to 'stack the deck' in its favor" before voir dire and trial.³ For example, in the seminal fair cross-section case *Duren v. Missouri*, women predictably comprised about half the relevant community but over time made up only about 15% of criminal venires, the relatively large panels from which trial or petit jurors are selected.⁴ Recognizing that the deck was badly and unjustifiably stacked against drawing female jurors, *Duren* held that the defendant was entitled to a fresh deal—at a new trial.⁵ Today, underrepresentation of distinctive groups is typically far less marked, leaving courts unsure of when a deck of jurors is so stacked as to give rise to a prima facie fair cross-section violation.

* J.D., Yale Law School, 2008; M. Phil., Cambridge University, 2005.

¹ A defendant establishes a prima facie fair cross-section violation by showing that (i) a "distinctive" group (ii) is not fairly and reasonably represented in venires over time (iii) due to "systematic exclusion" in the jury selection system (that is, in the jurisdiction's process of populating venires). *Duren v. Missouri*, 439 U.S. 357, 364 (1979). Once a defendant satisfies this three-pronged test, the government has the burden of justifying the challenged system. *Id.* at 368. Thus, a finding of unfair and unreasonable representation under the second *Duren* prong does not in itself guarantee the success of a fair cross-section claim: the third prong concerning systematic exclusion must also be satisfied, and the government must then fail to meet its burden of justifying the disparity. Importantly, fair cross-section claims do not require even the possibility of invidious intent, and so may be raised in response to, for example, computer malfunctions. *See, e.g., United States v. Jackman*, 46 F.3d 1240, 1242–43, 1246 (2d Cir. 1995). Intentional discrimination in jury selection, whether demonstrated directly or inferentially, is separately addressed through Equal Protection Clause jurisprudence. *See generally Castaneda v. Partida*, 430 U.S. 482 (1977).

² *See, e.g., Duren*, 439 U.S. at 364 ("[W]omen 'are sufficiently numerous and distinct from men' so that 'if they are systematically eliminated from jury panels, the Sixth Amendment's fair-cross section requirement cannot be satisfied.'" (quoting *Taylor v. Louisiana*, 419 U.S. 522, 531 (1975))).

³ *Holland v. Illinois*, 493 U.S. 474, 481 (1990).

⁴ *Duren*, 439 U.S. at 365–66.

⁵ *Id.* at 370.

The Supreme Court grappled with this abiding dilemma during the recent oral argument in *Berghuis v. Smith*.⁶ At one juncture, Justice Sonia Sotomayor suggested that a fair cross-section violation might result whenever a group comprising about 9% of the overall population is totally absent from venires.⁷ She then went on to acknowledge that if the relevant group were smaller, say “1 percent of the population, [then] it’s not likely that their absence is going to give rise to any flags.” “I think there is a difference” between the 9% and 1% scenarios, Justice Sotomayor continued, but “I just don’t know statistically where” to draw the line.⁸

Enter Justice Stephen Breyer, who had earlier attempted to shed light on the matter with his own analogy. “[T]he only way you could figure out . . . what’s what here is you use something called [the] ‘binomial theorem,’” Breyer explained.⁹

[Y]ou have to have, like, urns, and you imagine that there’s an urn with 1,000 balls, and 60 of them are red, and 940 are black, and then you select them at random, and—and 12 at a time. You know, fill 12—fill a hundred with 12 in each.¹⁰

This cryptic invocation of the binomial theorem prompted blank looks, and a hasty change of topic.¹¹

Justice Breyer had the right idea, but an unhelpfully obscure analogy. The following Commentary shows that statistical methods useful to poker players can and should inform our understanding of when venires are so unrepresentative as to violate the Sixth Amendment. Though I have elsewhere argued that the fair cross-section requirement should protect eligible jurors’ ability to participate in democratic self-governance through jury service,¹² I do not renew that argument here. Instead, the following Commentary accepts and takes seriously the conventional position, enshrined in Supreme Court precedent, that systematically underrepresentative venires harm criminal defendants by “stacking the deck”—that is, by reducing defendants’ ex ante odds of having “distinctive” group members included in their petit juries.¹³ As Justice Breyer recognized, that probabilistic

⁶ Transcript of Oral Argument, *Berghuis v. Smith*, 130 S. Ct. 1382 (2010) (No. 08-1402).

⁷ *Id.* at 10.

⁸ *Id.* at 12.

⁹ *Id.* at 5.

¹⁰ *Id.*

¹¹ *Id.* at 5–6.

¹² See Richard M. Re, Note, *Re-Justifying the Fair Cross Section Requirement: Equal Representation and Enfranchisement in the American Criminal Jury*, 116 *YALE L.J.* 1568 (2007); see also *Taylor v. Louisiana*, 419 U.S. 522, 530 (1975) (“Community participation in the administration of the criminal law” is “consistent with our democratic heritage.”).

¹³ See *Holland v. Illinois*, 493 U.S. 474, 477–78, 480–81 (1990).

injury, like the injury a poker player experiences when playing with a stacked deck, can be measured only with the help of the binomial theorem.¹⁴ And only after this harm is accurately measured is it possible to answer Justice Sotomayor's normative question of where to draw the constitutional line.¹⁵

This Commentary proceeds in four Parts. Part I develops the comparison between jury selection and poker in order to explain the disparity-of-risk test, which measures the probabilistic injuries associated with fair cross-section violations.¹⁶ Part II then proposes and defends a new 50% threshold for unfair and unreasonable representation. Next, Part III criticizes existing metrics of substantial underrepresentation—the absolute and comparative disparity tests—and provides courts with two judicially manageable means of employing the relatively complex disparity-of-risk test with a 50% threshold. Changing gears, Part IV argues that the standard deviation test should be understood as a test of “systematic exclusion” under *Duren*'s third prong, and not as a test of substantial underrepresentation under *Duren*'s second prong.¹⁷ Finally, the Conclusion suggests that the Supreme Court's recent decision in *Berghuis v. Smith*¹⁸ may augur a more statistically sophisticated approach to the fair cross-section requirement.

I. HOW TO MEASURE THE INJURY OF PLAYING WITH A STACKED DECK

Jury selection is a lot like poker. The queue of eligible jurors available for trial is analogous to a casino's deck of cards. And the use of peremptory and for-cause strikes during voir dire is like the card-trading phase of certain versions of poker, during which players attempt to influence the composition of otherwise randomly drawn hands.¹⁹ As the Supreme Court has recognized, the consequences of strategic decisions during voir dire do not shed light on whether the deck of eligible jurors was stacked at the outset.²⁰ Rather, to determine whether the deck is

¹⁴ See Transcript of Oral Argument, *supra* note 6, at 5.

¹⁵ See *id.* at 12.

¹⁶ *United States v. Jackman*, 46 F.3d 1240, 1254 (2d Cir. 1995) (Walker, J., dissenting) (discussing the disparity-of-risk test).

¹⁷ *Duren v. Missouri*, 439 U.S. 357, 364 (1979).

¹⁸ *Berghuis v. Smith*, 130 S. Ct. 1382, 1395–96 (2010).

¹⁹ Peremptory strikes normally allow parties to exclude eligible jurors without providing any explanation; for-cause strikes require a lawful justification. See generally *Batson v. Kentucky*, 476 U.S. 79 (1986); *Swain v. Alabama*, 380 U.S. 202 (1965).

²⁰ In *Holland v. Illinois*, 493 U.S. 474 (1990), the Supreme Court stated:

[T]o say that the Sixth Amendment deprives the State of the ability to “stack the deck” in its favor is not to say that each side may not, once a fair hand is dealt, use peremptory challenges to eliminate prospective jurors belonging to groups it believes would unduly favor the other side.

Id. at 481. In this way, the “fair-cross-section venire requirement assures . . . that in the process of selecting the petit jury the prosecution and defense will compete on an equal basis.” *Id.* *Holland* held that invidiously discriminatory use of peremptory strikes does not implicate the fair cross-section requirement, even though such actions plainly affect the defendant's ultimate “hand” of

stacked one must consider defendants' ex ante chances of being "dealt" particular twelve-juror hands, before or without voir dire.²¹

This is where the binomial theorem comes in. When the same probabilistic event—like flipping a coin or drawing a juror—occurs a specified number of times, the likelihood of any given number of results can be determined using the binomial theorem. In a simple case, the result can easily be acquired by hand. Consider, for example, the probability of being dealt no diamonds in two successive, independent deals, such as when the first card dealt is reshuffled into the deck before the second card is dealt.²² A standard 52-card deck has 13 cards of each suit. Thus, a standard deck has 39 non-diamonds, and the likelihood of not being dealt a diamond in any given one-card deal is 39/52. Squaring that figure supplies the likelihood of receiving a non-diamond in *both* of two independent deals. The likelihood of drawing zero diamonds after two independent one-card deals is thus $(39/52)^2$, or about 56%. Now imagine that four diamonds are dropped from the deck. The number of non-diamonds remains 39, but the total size of the deck has shrunk from 52 to 48. As a result, the risk of not being dealt a diamond on any given deal rises from 39/52 to 39/48, and the risk of not being dealt a diamond after two one-card deals rises to $(39/48)^2$, or about 66%. Thus, stacking the deck has increased the likelihood of drawing no diamonds by about 10%, from 56% to 66%.²³

The above example illustrates that if a card player hopes to draw cards of a particular suit, then the probabilistic injury of playing with a stacked deck depends on: (i) the odds of drawing a suited card from a standard deck; (ii) the odds of drawing a suited card from the stacked deck; and (iii) the total number of cards comprising the player's hand.²⁴

jurors. *Id.* at 481–84. Because of *Holland*, only Equal Protection Clause cases like *Batson*, 476 U.S. 79, police the constitutionality of "card-trading" decisions.

The fair cross-section requirement, by contrast, focuses on group representation in venires over time in order to ascertain whether the "deck"—that is, the process of populating venires—is unconstitutionally stacked, such that the composition of the defendant's initial "hand" has been unfairly affected. *Holland*, 493 U.S. at 481. A defendant who raises a successful fair cross-section challenge has defended not only his own interests, but also the interests of other defendants. All defendants in a given jurisdiction, after all, are dealt their initial hands from the same deck.

²¹ Viewed another way, the following analysis homes in on the effects of the jury selection system (as opposed to the effects of peremptory and for-cause strikes) by focusing on the first twelve eligible jurors put before the parties during voir dire. Unless otherwise specified, all calculations in this Commentary assume trial or petit juries composed of twelve jurors. However, other jury sizes are constitutional and sometimes used. See *Ballew v. Georgia*, 435 U.S. 223, 229–30 (1978).

²² The main text imagines reshuffling the first card back into the deck—and, later, that the cards are drawn from a very large deck—in order to ensure that each deal is "independent," that is, that the odds of drawing a diamond remain constant with each successive draw. Jurors are "dealt" independently; because the pool of eligible jurors is large, drawing one distinctive group member does not appreciably change the odds of drawing another.

²³ The main text statement is equivalent to saying that the likelihood of drawing at least one diamond has *diminished* by 10%.

²⁴ The third consideration—the size of the player's hand—is entirely overlooked by the

Now imagine a slightly more complicated scenario. A professional card player has joined an unusual poker game at a casino. The rules require that each player draw a 12-card hand from a very large deck that should contain 25% diamonds. After playing and counting cards at the table for a long time, the pro realizes that her hands on average contain only 15% diamonds. Because the actual deck is not behaving like a standard deck, the card player will infer that the deck is stacked, such that it contains only 15% diamonds and not the 25% that it ought to. This scenario is analogous to a jurisdiction in which petit juries include 12 people, a distinctive group comprises 25% of the overall population, and the distinctive group over time comprises only 15% of venirees. Like the repeat-playing, card-counting pro, someone managing jury selection in such a jurisdiction would conclude that something is wrong. A jury selection system should over time generate venirees that mirror the relevant jurisdiction's population of eligible jurors. Because the imagined jury selection system does not do that, the system is "stacked" and, if practicable, should be fixed. But how much are defendants really hurt by the skewed jury selection system?

To answer this question, return to the casino. Imagine that a second card player comes up to the same table described above to play the same unusual card game. Unlike the first card player, the second one is a dilettante and only intends to play one hand. We already know that the new player is about to draw cards from a stacked deck. The game is therefore unfair, and the dilettante's chances of drawing diamonds will be adversely affected. To measure the magnitude of this injury, we need to compare the odds facing the card player if the deck were fair with the actual odds facing the card player given that the deck is stacked. To make this task easier, we can use a binomial calculator.²⁵ Assume again that a hand is 12

statistical tests currently employed in fair cross-section cases. *See infra* Part II.

²⁵ A binomial calculator applies the binomial theorem:

$$(x + y)^n = \sum_{k=0}^n \binom{n}{k} x^k y^{n-k}$$

When x is defined as the probability of a "positive" result and y as the probability of a "negative" result, the binomial theorem defines a binomial distribution. The term x^k measures the odds of k "positive" results, the term y^{n-k} measures the odds of $(n-k)$ "negative" results, and $\binom{n}{k}$ measures the number of different ways that one can obtain k positive results and $(n-k)$ negative results. The function $\binom{n}{k} x^k y^{n-k}$ thus measures the odds of exactly k positive results after n attempts. To identify the odds of obtaining a *range* of positive rules—such as all positive results between k and $(k + 2)$ —simply perform the $\binom{n}{k} x^k y^{n-k}$ function for each k value within the desired range and sum the resulting probabilities. If the $\binom{n}{k} x^k y^{n-k}$ function is performed for each and every possible number of positive results from zero to n , the resulting probabilities will sum up to 1, or 100%. This fact is captured in the binomial theorem by the $\sum_{k=0}^n$ summation symbol. (In the example from the main text that immediately follows this footnote, n is 12, x is 0.25, y is 0.75, and k ranges from 0 to 2.) More

cards, that the deck should include 25% diamonds, and that the actual observed likelihood of a diamond on any given deal is 15%. In this example, the expected number of diamonds is 3 (that is, 12 times 0.25), and a binomial calculator reveals that the odds of being dealt fewer than 3 diamonds are about 39%. Of course, this figure does not mean that there is anything wrong with the deck; on the contrary, the deck by hypothesis is perfectly standard. The 39% chance of a below-expectation deal simply demonstrates that one does not always get what one expects when playing cards at a casino—or, for that matter, when drawing jurors for trial. Luck intervenes.

Now we have to determine the actual odds facing the card player given that the deck is stacked. Just as with the standard deck, a card player unwittingly playing with a stacked deck will expect 3 diamonds. The card player's expectations, in other words, remain the same whether the deck is standard or stacked. Only the player's actual odds of obtaining a below-expectation deal will change. Given that a hand is still 12 cards and that the odds of drawing a single diamond from the stacked deck are 15%, the odds of receiving a below-expectation hand (that is, of being dealt 0, 1, or 2 diamonds) rise to 73.5%. In other words, when the dilettante stepped up to the table to play a single hand, he reasonably foresaw a 39% chance of a below-expectation deal. In fact, however, the stacked deck raised the chance of that misfortune all the way to 73.5%. What was once unlikely has thus become likely. This 34.5% swing in absolute expected probabilities is an appreciable injury to the dilettante's card-playing chances—just as it would be to a criminal defendant who hopes to avoid an underrepresentative number of distinctive group members on his jury.

The above analytic approach is called the “disparity-of-risk analysis,” and its inventor, Peter A. Detre, provided the following summary of the procedure:

detailed explanations are available in any basic statistics book. See, e.g., MICHAEL O. FINKELSTEIN & BRUCE LEVIN, *STATISTICS FOR LAWYERS* 103–06 (2d ed. 2001).

Because readers may be unfamiliar with some of the mathematical symbols used in the binomial theorem, it is worthwhile to explain what each symbol means.

Start with the summation symbol, $\sum_{k=0}^n$. This term is read “the sum from zero to n ” and means that the operation that follows the summation symbol is executed using as k all integers from k to n . That is, the subsequently defined operation is performed first with k as zero, then with k as one, then with k as two, etc. until k is equal to n , at which point all the previously identified products are added together. For example, $\sum_1^2 k$ would mean “the sum of the operation ‘ k ’ from $k = 1$ to $k = 2$,” or $(1 + 2)$.

Next, take the symbol, $\binom{n}{k}$. This term is read “ n choose k .” It measures the number of ways that k objects can be drawn from a pool of n objects and entails the following operation: $\frac{n!}{k!(n-k)!}$.

The symbol ! is read as “factorial” and means that all integers from n down to 1 are multiplied by one another. For example, $4!$ or “four factorial” in effect means “four times three times two times one,” or $(4*3*2*1) = 24$. And $\binom{5}{2}$ or “five choose 2” means $\frac{5!}{2!(5-2)!} = \frac{5*4*3*2}{(2)(3*2)} = 10$.

In general, to determine the effect that the underrepresentation of a given group has on the defendant's ex ante chances of an unrepresentative jury, calculate the disparity of risk as follows. First determine the expected number of members of the group in question per jury if the drawing were done at random from a representative wheel [EV]. For each integer, $[k]$, strictly less than this expected number, calculate the probability that a jury would have $[k]$ or fewer group members on it, first, if the wheel were representative $[p_1]$, and second, given the actual underrepresentation $[p_2]$. For each $[k]$, of course, the second probability will be larger. The difference between these two probabilities is a measure of the change in the defendant's chances of drawing a jury with $[k]$ or fewer group members on it. The greatest probability increase as $[k]$ runs from zero up to (but not including) the expected number of group members $[P(p_1) - P(p_2)]$ is defined to be the defendant's disparity of risk. In effect, disparity of risk measures the amount by which underrepresentation on the wheel increases the defendant's risk of drawing an unrepresentative jury.²⁶

With the aid of a binomial calculator (freely accessible online), implementing the disparity-of-risk test is far easier than it looks.²⁷ One need only identify the number of jurors per panel (n), the distinctive group's percentile portion of the overall population (p_1), the distinctive group's actual representation in venires over time (p_2), and the set of integers less than the expected value (EV) of distinctive group members (k). (Don't worry: Part III will provide ways of using statistical insights without any calculation at all.) To better illustrate this process, the following table shows the relevant figures for the above example (12-card hand where the odds of a diamond should be 25% but are actually 15%):

Table 1

Distinctive Group Members Drawn k	Ideal Risk with a Fair Deck $P(p_1 = .25)$	Actual Risk with Stacked Deck $P(p_2 = .15)$	Disparity of Risk $P(p_2) - P(p_1)$
≤ 2	.391	.736	.345
≤ 1	.158	.444	.286
$= 0$.032	.142	.110

²⁶ Peter A. Detre, Note, *A Proposal for Measuring Underrepresentation in the Composition of the Jury Wheel*, 103 YALE L.J. 1913, 1932–33 (1994) (citations omitted) (alterations not in original).

²⁷ Even easier, a simple computer program could conveniently execute these operations and yield the largest disparity of risk figure (whereas the binomial calculator is useful only for filing in individuals cells in the table above).

The columns “ $P(p_1)$ ” and “ $P(p_2)$ ” respectively indicate the risk of having k or fewer distinctive group members when the jury selection process is perfectly representative (p_1) and when it is skewed (p_2), while the “ $P(p_2) - P(p_1)$ ” column lists the resulting disparity-of-risk values.

Importantly, disparity-of-risk values must be separately ascertained for each k value. This additional work is necessary because stacking a deck can simultaneously upset numerous expectations, including the expectation to draw any given number of distinctive group members fewer than the expected value. In the table above, for example, the increased risk of having two or fewer distinctive group members is larger than the increased risk of having either one or fewer distinctive group members or zero distinctive group members. The largest disparity of risk—here, the bolded 34.5% figure—identifies and measures the defendant’s ex ante expectation that is most disrupted by the fact that the deck has been stacked. If that probabilistic injury is deemed legally substantial, then the defendant would have satisfied *Duren*’s second prong. But if that probabilistic injury is deemed legally de minimis, then courts might plausibly discount all the other, lesser probabilistic injuries suffered by the defendant. In other words, the largest disparity-of-risk figure represents the defendant’s strongest claim to a legally cognizable probabilistic injury.²⁸ Having arrived at that statistical measurement, we are now ready to ask how large the probabilistic injury must be to sustain a fair cross-section claim.

II. A PROPOSED THRESHOLD FOR UNFAIR AND UNREASONABLE UNDERREPRESENTATION

Mathematics can measure the probabilistic effects of distorted jury selection procedures. But mathematics cannot tell us when those effects are so large as to make the composition of venires no longer “fair and reasonable.”²⁹ Detre

²⁸ The decision to focus legal attention on a defendant’s single most disrupted expectation rests significantly on a value choice. A critic might argue, for example, that a defendant’s single most disrupted expectation should be less legally important than the defendant’s expectation of having at least one distinctive group member on his petit jury, or of having at least some larger “critical mass” of distinctive jurors. But by focusing on the single largest disparity-of-risk, the approach adopted in the main text errs on the side of the defendant by permitting the defendant to rely on whatever single expectation is most affected. Alternatively, one might argue that courts should focus on the average of multiple disrupted expectations, such that a sufficiently large number of individually de minimis diminished expectations might give rise to a fair cross-section violation. However, an average-based approach would sometimes cut against defendants. And as Detre pointed out: “[I]f the probability of a certain kind of unrepresentative jury increases substantially, then the defendant is exposed to that additional risk [and] there does not appear to be any reason to downgrade it simply because the increased risk of a different sort of injury is less dramatic.” Detre, *supra* note 26, at 1933 (emphasis omitted). In any event, the appeal of the average-based approach and other alternatives depends in large part on what normative threshold is chosen to divide legally substantial and legally de minimis probabilistic injuries. See *infra* Part II.

²⁹ *Duren v. Missouri*, 439 U.S. 357, 364 (1979).

recognized this limitation.³⁰ So to ascertain an appropriate standard for fair and reasonable representation, Detre reasoned from what he considered to be an uncontroversial case. In particular, Detre assumed that a fair cross-section violation obtains when a group comprises 50% of the overall population but only 40% of grand juries over time.³¹ Detre then calculated the disparity of risk for that scenario and set the resulting figure—37%—as his suggested threshold level for fair and reasonable representation.³²

Though Detre invaluablely contributed to the statistical analysis of jury underrepresentation, his proposed 37% threshold is problematic. Again, Detre arrived at that figure by reasoning from the assumption that an arbitrarily selected scenario yielded a fair cross-section violation. Detre's only reason for selecting the scenario that he did—a 50% population group underrepresented on 23-member grand juries by 10%—was his belief that the scenario described a degree of underrepresentation that most or all courts of appeals would deem actionable.³³ Yet the same could be said of many other scenarios involving different population percentages, disparities, and panel sizes. And even a relatively small change in any of these variables could yield a significantly different disparity of risk. For example, if Detre had posited that the archetypical fair cross-section challenge occurred when a distinctive group comprised 40% of the population and was underrepresented in venires by 10%, he would have arrived at a disparity-of-risk threshold of about 30% instead of 37%. Detre offers no reason to prefer either of these figures, or any number of other potential figures.

So what is the appropriate legal threshold beyond which a disparity of risk becomes unfair and unreasonable? Before attempting to answer that question, it is important to repeat that choosing a threshold for unfair and unreasonable representation constitutes a legal or normative decision, and is not a mathematical or descriptive one. The disparity-of-risk test can tell us when a defendant's probabilistic injury is more or less severe, but it cannot and does not tell us when such an injury should be legally cognizable. There is no line beyond which a probabilistic harm suddenly becomes mathematically "substantial," where before it was only trivial. That undeniable fact may argue against choosing a hard-and-fast threshold at all. Instead, probabilistic injuries might simply be viewed as falling along a spectrum. Yet the problem of whether and where to draw statistically arbitrary lines is not unique to the fair cross-section requirement. On the contrary, probabilistic judgments pervade the law, if not all practical decision-making.³⁴

³⁰ See Detre, *supra* note 26, at 1936–37.

³¹ *Id.* at 1936. Though Detre does not explicitly assume a grand jury pool, his stated conclusion mathematically depends on an *n* value of 23, consistent with his repeated use of grand jury examples. See *id.* at 1929 & n.75, 1934–35, 1935 & n.90.

³² See *id.* at 1936–37.

³³ See *id.*

³⁴ See generally FREDERICK SCHAUER, PROFILES, PROBABILITIES, AND STEREOTYPES (2003) (discussing the validity and injustice of stereotyping and profiling); see also text accompanying *infra*

The fair cross-section doctrine can therefore be seen as an opportunity to apply established legal principles in a particular context. Viewed in that light, disparities of risk might reasonably be deemed unfair and unreasonable only when they exceed 50%.

To see the 50% threshold's normative appeal, contrast two hypothetical scenarios. The first scenario involves a distinctive group that makes up 5% of the overall population but 0% of venirees. If the imagined jury selection system had succeeded in dealing distinctive group members 5% of the time, then the risk of having zero distinctive group members on any given petit jury would be about 54%. But because the distinctive group is totally absent from venirees, the actual risk of drawing no distinctive group members rises to 100%, yielding a disparity of risk of (100% – 54%), or 46%. This 46% disparity-of-risk figure captures the fact that defendants were unlikely to have any distinctive group members in their petit juries even if the jury selection system had worked perfectly. Because the defective jury selection system did not cost defendants anything that they were more likely than not to receive, one might plausibly conclude that no defendant was legally harmed. Such a conclusion would parallel the commonplace legal rule that claimants are entitled to no relief when they fail to show it is more likely than not that they have been wronged.³⁵

Now consider the second hypothetical scenario: a distinctive group that comprises 6% of the overall population and, again, 0% of venirees. If *that* jury selection system had succeeded in dealing distinctive group members 6% of the time, then the risk of having zero distinctive group members on any given petit jury would be about 47%. But because the distinctive group is totally absent from venirees, the actual risk of drawing no distinctive group members rises to 100%, for a disparity-of-risk score of (100% – 47%), or 53%. In this hypothetical, unlike the first one, each defendant would indeed have lost something that he reasonably expected to obtain—namely, at least one distinctive group member on his petit jury. Applying the 50% threshold for disparity of risk in this scenario reflects the fact that the jury selection system's defects caused the eradication of a reasonable (because more-likely-than-not) expectation.³⁶ The 50% threshold parallels the

note 57.

³⁵ Preponderance standards are ubiquitous, governing for example most tort, contract, and evidentiary inquiries. See SCHAUER, PROFILES, PROBABILITIES, AND STEREOTYPES 89, 315–16 nn.9–11. This may seem odd “to a statistician,” given that:

[T]he law would give Smith all of her damages if she proved her case to a .51 probability, and nothing if she proved it to a .49 probability. And it would give her not a dollar more if she proved her case to a .90 probability than if she proved it to a .51 probability. . . . [Advanced legal] systems, we see throughout the world, are all-or-nothing affairs.

Id. at 89. In special contexts, of course, the preponderance default rule is set aside, such when criminal prosecutions rest on a beyond-a-reasonable-doubt showing. See *id.*

³⁶ Do not confuse reasonable, more-likely-than-not expectations with the “expected value.” Cf. *United States v. Royal*, 174 F.3d 1, 8 & n.5 (1st Cir. 1999) (discussing the “absolute impact” test, also known as the “absolute numbers” test, whereby courts effectively calculate the change in the

standard legal rule that claimants are entitled to complete recovery even when they only barely satisfy a preponderance-of-the-evidence causation standard. And eliminating the 50% disparity of risk in this scenario would, more likely than not, change the composition of the defendant's jury.³⁷

The disparity-of-risk test allows courts to generalize the normative insight acquired by contrasting the two foregoing scenarios. For example, the 50% disparity-of-risk threshold is also satisfied when a group comprises 25% of the overall population but no more than 10% of venires. To be sure, the risk of drawing an underrepresentative jury does not rise to 100% in that case, as it did in the preceding scenario. But any given defendant's risk of drawing an underrepresentative jury—in particular, her risk of drawing one or fewer distinctive group members—does rise to a mathematically comparable extent, from 16% to 66%. In other words, the jury selection system's defects were sufficiently severe that they alone increased the odds of drawing an underrepresentative jury by 50%. Therefore, the jury system's defects adversely affected the defendant's reasonable expectations to a legally substantial degree.

As the two contrasted examples above illustrate, the 50% threshold, when applied to 12-person petit juries, has the effect of ruling out fair cross-section claims based on distinctive groups comprising less than about 5.7% of the total population.³⁸ This result is consistent with Justice Sotomayor's view that there is a normatively meaningful difference between the total absence of groups comprising 1% and 9% of the population.³⁹ The 50% disparity-of-risk threshold thus has the important advantage of being consistent with commonsense intuitions about how the fair cross-section requirement should operate in practice, particularly in those especially difficult and frequent cases involving small distinctive groups.

Despite the foregoing arguments, some skeptics may reasonably insist that the 50% threshold imposes too high of a bar on criminal defendants. A critic might

expected value resulting from the group's underrepresentation in venires); *United States v. Jenkins*, 496 F.2d 57, 64–66 (2d Cir. 1974) (discussing and applying the “absolute numbers” test). The “expected value” in this context is the average number of distinctive group members in trial juries. In the main text example associated with this footnote, the expected value is 12 times 0.06, or 0.72 distinctive group members. That decimal figure indicates that, on average, each petit jury will include less than one distinctive group member. However, no humane criminal justice system slices jurors into fractions. Given that real-world constraint, the most likely draw is one distinctive group member, as indicated by the use of the binomial theorem in the main text.

³⁷ *Cf. supra* note 35. Probabilistic notions of redressability are also visible in Article III standing doctrine. *See Summers v. Earth Island Inst.*, 129 S. Ct. 1142, 1149–50 (2009) (“[I]t must be likely that a favorable judicial decision will prevent or redress the injury.”); *Vt. Agency of Natural Res. v. U.S. ex rel. Stevens*, 529 U.S. 765, 771 (2000) (The plaintiff “must demonstrate redressability—a ‘substantial likelihood’ that the requested relief will remedy the alleged injury in fact.” (quoting *Simon v. E. Ky. Welfare Rights Org.*, 426 U.S. 26, 45 (1976))).

³⁸ The total exclusion of a 3% population group from 23-member grand juries would also satisfy the 50% disparity-of-risk standard. Throughout this Commentary, calculations assume twelve-person petit juries, unless grand juries are specified.

³⁹ *See supra* text accompanying notes 6–11.

argue, for example, that a card player would properly refuse to honor her bets after learning that her odds of drawing certain desirable hands had been diminished by less than 50%.⁴⁰ Such analogies may not be dispositive, as our normative intuitions regarding fair play in games of chance do not necessarily extend to the more administratively and morally complex context of fair cross-section litigation.⁴¹ In any event, the 50% threshold has something to offer even determined proponents of a lower bar—namely, a bright-line *sufficient* condition for unfair and unreasonable underrepresentation. In other words, one might conclude that any defendant who establishes a 50% disparity of risk necessarily establishes unfair and unreasonable underrepresentation, even if a less impressive disparity-of-risk figure might suffice under some circumstances. That more modest conclusion would still have significant practical implications. As the next Part will show, using the 50% threshold as a sufficient condition for establishing unfair and unreasonable representation would permit many fair cross-section claims foreclosed under the dominant approach employed by the federal courts of appeals.

III. CRITIQUING AND IMPROVING EXISTING MEASUREMENTS OF UNDERREPRESENTATION

Instead of employing the disparity-of-risk test, courts have looked almost exclusively to two other tests—the absolute disparity test and the comparative disparity test—when measuring substantial underrepresentation in criminal venires.⁴² In *Berghuis v. Smith*, the Supreme Court recently called these tests “imperfect” and potentially “misleading,” particularly when relatively small groups are at issue.⁴³ The Court’s skepticism is warranted. Although each of these established tests captures an important consideration that the other overlooks, neither test addresses the legally relevant issue: the reduction in an individual defendant’s *ex ante* likelihood of drawing a particular set of jurors.⁴⁴

Under the leading test, called the absolute disparity test, a group’s underrepresentation is measured by subtracting its percentile representation in venires from its percentile representation in the overall population.⁴⁵ For example,

⁴⁰ Cf. *Detre*, *supra* note 26, at 1929 (“Anyone who has ever bet on a horse race would have to regard [an approximately 20%] change in odds as ‘substantial.’”). Despite this statement, *Detre* ultimately endorsed a 37% threshold. See *supra* text accompanying note 32.

⁴¹ Cf. *Hamling v. United States*, 418 U.S. 87, 138 (1974) (“[S]ome play in the joints of the jury-selection process is necessary in order to accommodate the practical problems of judicial administration.”).

⁴² See *Berghuis v. Smith*, 130 S. Ct. 1382, 1393 (2010).

⁴³ *Id.*

⁴⁴ See generally *Detre*, *supra* note 26, at 1927–30 (explaining that the established tests are simply “measuring the wrong thing”).

⁴⁵ See, e.g., *United States v. Royal*, 174 F.3d 1, 10 (1st Cir. 1999); *Thomas v. Borg*, 159 F.3d 1147, 1150 (9th Cir. 1998); *United States v. Rioux*, 97 F.3d 648, 655–56 (2d Cir. 1996); *United*

when a group comprises 50% of the overall population and 25% of venirees, the absolute disparity is (50% – 25%) or 25%. Returning to the deck-of-cards analogy, the absolute disparity test in effect provides a percentile measurement of how many suited cards are dropped from the deck, relative to the total number of cards in the deck. The absolute disparity test in the foregoing example thus answers the question, “How many diamonds were dropped?” by saying, “A quarter of the deck.” But if a card player is looking for diamonds, then the effect of the dropped diamonds is not solely determined by the ratio of dropped diamonds to total cards. The effect is also determined in part by how many diamonds remain in what may be a nonstandard deck. Analogously, knowing that a jurisdiction exhibits an absolute disparity of, say, 50% does not reveal defendants’ ex ante likelihood of drawing distinctive group members for their petit juries. If the jurisdiction is 99% African American and venirees are 49% African American, then defendants would be virtually assured of having African Americans on their petit juries, despite the 50% absolute disparity.⁴⁶ If, on the other hand, the overall population is 50% African American and venirees are 0% African American, then the odds of having an African American petit juror would drop from near-certainty to total impossibility. The fact that the absolute disparity test cannot distinguish between these radically different scenarios indicates that it does not measure defendants’ probabilistic injuries.

The comparative disparity test encounters the opposite fatal flaw. The comparative disparity test divides the absolute disparity figure by the distinctive group’s percentile representation in the overall population.⁴⁷ In doing so, the comparative disparity test in effect captures the ratio of dropped diamonds to remaining diamonds. Consider again a 50% population group that comprises 25% of venirees. Because that scenario involves an absolute disparity of 25% and a 50% population group, the resulting comparative disparity is (25% / 50%), or 50%. This figure answers the question, “How many diamonds were dropped?” by saying, “Half of the diamonds.” Because it considers the fraction of diamonds that remain in the deck, the comparative disparity test fills its absolute counterpart’s most conspicuous blind spot. Yet the comparative disparity test lacks the absolute

States v. Rogers, 73 F.3d 774, 775–77 (8th Cir. 1996); United States v. Ashley, 54 F.3d 311, 313–14 (7th Cir. 1995). Courts generally require an absolute disparity of 10% before finding a violation. See, e.g., Ashley, 54 F.3d at 314. This prevailing approach has drawn extensive criticism because it forecloses claims based on groups comprising less than 10% of the total population. See, e.g., United States v. Jackman, 46 F.3d 1240, 1246–48 (2d Cir. 1995).

⁴⁶ Readers who still find this hypothetical objectionable (because it suggests that half of all eligible jurors have been excluded from venirees over time) may prefer an enfranchisement-based understanding of the fair cross-section requirement. See *supra* note 12 and accompanying text.

⁴⁷ Courts that consider the comparative disparity test have found disparities as high as about 60% to be inadequate to support a *prima facie* case. See, e.g., United States v. Orange, 447 F.3d 792, 798–99 (10th Cir. 2006) (discussing United States v. Shinault, 147 F.3d 1266, 1272–73 (10th Cir. 1998)). The comparative disparity test is often criticized for overstating the underrepresentation of small groups. See, e.g., United States v. Hafén, 726 F.2d 21, 24 (1st Cir. 1984). For example, the total exclusion of any group, no matter how small, yields a comparative disparity of 100%. *Id.*

disparity test's awareness of what fraction of the total deck has been tampered with. For example, when all African Americans are absent from venires, the result is the highest possible comparative disparity score, 100%. But that figure is useless unless one also accounts for how many African Americans are in the overall population. If the total population is majority African American, then the observed underrepresentation would reduce the odds of drawing an African American juryperson from near certainty to total impossibility. If, on the other hand, African Americans comprise just 0.1% of the total population, then the likelihood of drawing an African American would not have significantly declined. Thus, despite its support among prominent commentators,⁴⁸ the comparative disparity test, like the absolute disparity test, simply does not measure the probabilistic injuries generated by fair cross-section violations.

The fact that the absolute and comparative disparity tests have corresponding weaknesses means that they also have complementary strengths. Intuitively recognizing this fact, some courts have rejected exclusive use of either test in favor of using both tests.⁴⁹ However, these courts have not explained how or why the outcomes from these tests should be used together to produce principled results. The disparity-of-risk test supplies the answer, as the absolute and comparative disparity tests can be used in combination to approximate the disparity-of-risk test. Consider the following three simple rules, which roughly approximate the disparity-of-risk test with a 50% threshold when groups comprise 15% or less of the overall population: (i) absolute disparities of 5% or less are never actionable;

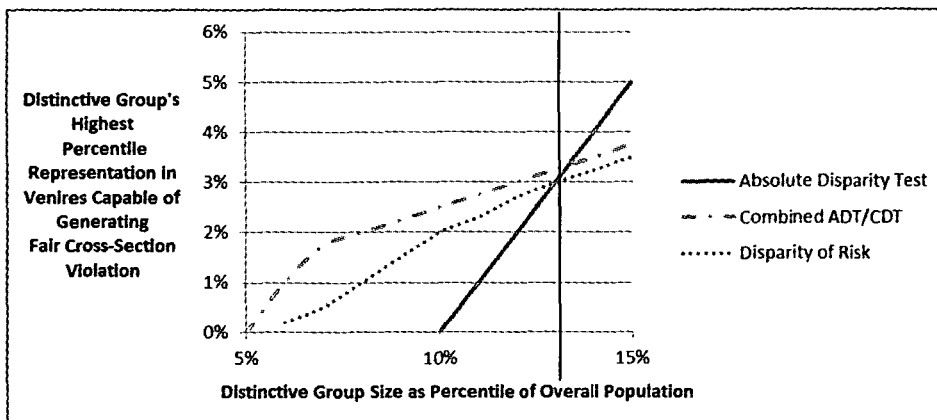
⁴⁸ More than two dozen academics submitted an amicus brief in *Berghuis v. Smith* arguing in favor of the comparative disparity test. See Brief for Social Scientists, Statisticians, and Law Professors, Jeffrey Fagan, et al., as *Amici Curiae* Supporting Respondent, *Berghuis v. Smith*, 130 S. Ct. 1382 (2010) (No. 08-1402) [hereinafter Brief for Social Scientists]. The brief asserted that the Supreme Court “should require 15 percent comparative disparity, except where the distinct group is less than 10 percent, in which case the Court should use 25 percent.” *Id.* at 33 (citing David Kairys, Joseph Kadane & John Lehoczky, *Jury Representativeness: A Mandate for Multiple Source Lists*, 65 CALIF. L. REV. 776, 796 n.112 (1977)). Neither the brief nor the article it cites supplies any reason to choose these particular thresholds, which would mean, for example, that a violation could occur when a 12% population group is represented in venires at 10%. *But see supra* note 47 (noting that those courts that consider the comparative disparity test find comparative disparities as high as about 60% to be inadequate). Further, the brief's exclusive reliance on the comparative disparity test necessarily defies Justice Sotomayor's commonsensical and statistically well-founded point that the total exclusion of a 1% population group should not generate a fair cross-section violation. See *supra* text accompanying note 8. Anticipating this objection, the brief candidly acknowledged that the comparative disparity approach “can overstate the underrepresentation of very small groups.” Brief for Social Scientists, *supra*, at 33. And, in a footnote, the brief further conceded that, “at some point, the population proportion becomes so small that the results are of no constitutional consequence.” *Id.* at 33 n.23. But instead of explaining when the comparative disparity test ceases to be useful, the brief instead contended that the Court could dispose of the case at hand without grappling with the issue. *Id.* That view is mistaken. As the main text associated with this footnote explains, the comparative disparity test's failure to make sense when applied to small groups is just one symptom of its general inability to measure defendants' probabilistic injuries.

⁴⁹ See, e.g., *Orange*, 447 F.3d at 798–99; *United States v. Weaver*, 267 F.3d 231, 243 (3d Cir. 2001).

(ii) absolute disparities between 5% and 12% are actionable only if accompanied by a comparative disparity of at least 75%; and (iii) absolute disparities above 12% are always actionable. Because it employs tests already familiar to courts, this combined absolute/comparative disparity test offers a judicially manageable means of employing the statistical insights gleaned from the disparity-of-risk analysis.

The graph below illustrates the relationship between the absolute disparity test with a 10% threshold (which, again, is the leading approach), the disparity-of-risk test with a 50% threshold, and the combined absolute/comparative disparity test just proposed. Each line represents what might be called an “actionability horizon.” All values below each line are actionable for the test associated with that line. A lower line thus translates into a more stringent test. The three lines roughly intersect at $x = 13\%$, as highlighted by the vertical line. To the left of the vertical line, the absolute disparity test is the most stringent; to the right, the disparity-of-risk test is the most stringent. In practice, the disparity-of-risk test and the combined test are more favorable to criminal defendants than the absolute disparity test: not only does fair cross-section litigation frequently involve groups smaller than 13% of the overall population, but the absolute disparity test with a 10% threshold entirely rules out claims based on groups comprising less than 10% of the overall population. In contrast, any group that can generate a violation under the absolute disparity test can also generate a violation under either the disparity-of-risk test or the combined absolute/comparative test.

Figure 1: Graph of Approximate Actionability Horizons



For another way of using the disparity-of-risk test without the hassle of computing the relevant values, the table below lists what might be called approximate “actionability values” for the disparity-of-risk test with 50% threshold. An actionability value is the highest venire representation that, given a particular group’s overall population representation, satisfies the 50% disparity-of-

risk threshold. To take an arbitrary example from the table below, consider the two bolded cells—30% in the “Population %” column and 14% in the “Venires %” column. This pairing means that whenever a group comprises 30% of the overall population and 14% or less of venires over time, the 50% disparity-of-risk threshold is satisfied. So venire representation at 13% would also be actionable for a 30% population group, whereas venire representation at 15% would not. When confronted with percentages between the values supplied on the table, courts can err on the side of criminal defendants by rounding fractional values up to the nearest listed population percentile and down to the nearest listed venire percentile. In this way, courts can approximate the disparity-of-risk test and 50% threshold without having to engage in any calculation whatsoever.

Figure 2: Table of Approximate Actionability Values for the Disparity-of-Risk Test with 50% Threshold

Population %	Venires %	Population %	Venires %	Population %	Venires %
1%	-	19%	6.50%	37%	20%
2%	-	20%	7%	38%	20.5%
3%	-	21%	8%	39%	21%
4%	-	22%	8.50%	40%	22%
5%	-	23%	9%	41%	23%
6%	0%	24%	9.50%	42%	23.5%
7%	0.50%	25%	10%	43%	24%
8%	1%	26%	11%	44%	25%
9%	1.50%	27%	12%	45%	26%
10%	2%	28%	13%	46%	27%
11%	2.25%	29%	13.5%	47%	28%
12%	2.50%	30%	14%	48%	29%
13%	3%	31%	15%	49%	30%
14%	3.25%	32%	15.5%	50%	31%
15%	3.50%	33%	16%	51%	31.5%
16%	4%	34%	17%	52%	32%
17%	5%	35%	18%	53%	34%
18%	6%	36%	19%	54%	35%

IV. STATISTICAL SIGNIFICANCE AS A TEST OF SYSTEMATIC EXCLUSION

Besides the absolute and comparative disparity tests, the Supreme Court's *Berghuis v. Smith* decision noted one other potential metric for evaluating substantial underrepresentation: the sometimes mentioned but never relied on standard deviation test.⁵⁰ On reflection, however, the standard deviation test is better understood as a measure, not of substantial underrepresentation, but rather of "systematic exclusion."⁵¹

The standard deviation or statistical significance test ascertains the likelihood that random chance explains a particular set of observed results.⁵² The magnitude of a standard deviation (*SD*) can be ascertained in the jury selection context by taking the square root of the product of three terms: the number of observed venirepersons (*N*), the expected probability that any particular venireperson is a member of a distinctive group (*P*), and the expected probability that any particular venireperson is not a member of that group ($1 - P$). Or, put arithmetically: $SD = \sqrt{(N)(P)(1 - P)}$.⁵³ The resulting output is the number of venirepersons constituting one standard deviation. Having identified the value of a standard deviation, the standard deviation test then considers the disparity between the number of distinctive group members expected in and actually present in an observed number of venires. So, for example, one might expect that, on average,⁵⁴ 100 out of 500 venirepersons will be distinctive group members, but observe that the actual figure is only 50. The resulting disparity would be $(100 - 50)$, or 50 venirepersons. As that observed disparity becomes larger in relation to the standard deviation, the likelihood that the disparity is the product of random chance declines. In the jury selection context (and other situations involving a binomial distribution), it is reasonable to assume that if the observed disparity is two standard deviations in magnitude, then there is a 5% probability that the

⁵⁰ *Smith*, 130 S. Ct. at 1393 (quoting *Rioux*, 97 F.3d at 655) (noting that no court has relied on the standard deviation test in finding a fair cross-section violation); see generally Michael O. Finkelstein, *The Application of Statistical Decision Theory to the Jury Discrimination Cases*, 80 HARV. L. REV. 338 (1966) (describing, before the crystallization of fair cross-section doctrine, the usefulness of the statistical significance test for proving invidious discrimination in jury selection).

⁵¹ *Duren v. Missouri*, 439 U.S. 357, 364 (1979).

⁵² *Smith*, 130 S. Ct. at 1390 n.1 ("Standard deviation analysis seeks to determine the probability that the disparity between a group's jury-eligible population and the group's percentage in the qualified jury pool is attributable to random chance." (citing *People v. Smith*, 615 N.W.2d 1, 9-10 (Mich. 2000) (Cavanagh, J., concurring))).

⁵³ This formula, too, assumes a binomial distribution. See *supra* notes 22 and 25. As a rule of thumb, using this formula to apply the standard deviation test will produce meaningful results only when it is true both that *N* times *P* is at least 5 and that *N* times $(1 - P)$ is at least 5. See, e.g., FINKELSTEIN & LEVIN, *supra* note 25, at 115-16.

⁵⁴ For a discussion of expected values, see *supra* note 36.

observed disparity is the result of random chance; and if the observed disparity is three standard deviations, the same probability is only about 0.3%.⁵⁵

Importantly, the likelihood that random chance can explain any given percentile disparity between observed and expected results becomes smaller as the number of total observations increases. Imagine for example that P and $(1 - P)$ both equal 0.5. If N is 10, then SD equals $\sqrt{(10)(.5)(1-.5)}$, or about 1.5. But if N is 1,000, then SD equals $\sqrt{(1000)(.5)(1-.5)}$, or about 15. Thus, N increased by a factor of 100, but SD increased by a factor of just 10. These examples illustrate that the standard deviation (SD) will become a smaller percentage of N as N becomes larger. If N becomes sufficiently large, for example, even a 0.01% disparity between expected and observed results could constitute a standard deviation. Common sense tells us this is correct. If we flip a coin a few times, random chance could easily cause a streak of three heads in a row. But if we flip a coin a thousand times, we strongly expect to see heads close to 50% of the time. If after 1,000 flips we saw heads only 45% of the time, we would be confident that something besides random chance had intervened. The standard deviation test bears out that well-founded intuition.

So when is an observed disparity deemed statistically significant, that is, sufficiently *unlikely* to have been caused by random chance to be deemed reliable? Any bright-line criterion for statistical significance—much like the selection of a threshold for fair and reasonable representation—is arbitrary from the standpoint of mathematics. Under widespread scientific and social scientific conventions, however, results are generally considered to be “statistically significant” at two or more standard deviations.⁵⁶ Consistent with that practice, the Supreme Court’s Equal Protection Clause cases have suggested that a disparity of two or three standard deviations establishes significance in the jury selection context.⁵⁷

The problem with using the standard deviation test to measure fair and reasonable representation is that the test measures the non-randomness of a given disparity, but not the disparity’s substantiality. This critical point becomes clear when one considers the very real possibility of disparities that are non-random and yet *de minimis*. For example, given a sufficiently large sample, it may be shown to a statistical certainty that a jury selection defect has increased the risk of an unrepresentative jury by just 0.1%. Such a finding would indicate that something is awry and may justify corrective measures, if practicable. But such a finding hardly undermines the legitimacy of the jurisdiction’s criminal convictions during the relevant time period. Therefore, a court faced with such evidence would be justified in declining to reverse any defendant’s conviction. By analogy, one might statistically prove that an incompetent casino dealer accidentally drops one diamond every 1,000 deals. Of course, the house should fully abide by the rules,

⁵⁵ See, e.g., FINKELSTEIN & LEVIN, *supra* note 25, at 113.

⁵⁶ See, e.g., *id.* at 120.

⁵⁷ See *Castaneda v. Partida*, 430 U.S. 482, 496–97 n.17 (1977).

and even rare or inadvertent dealing errors should stop. But because no individual card player's chances have been substantially injured, the casino might fairly insist that all players honor their past bets, despite the dealing errors.

Still, the standard deviation test is relevant to fair cross-section claims. When a defendant cannot demonstrate that an observed disparity in jury venires is statistically significant, the defendant's statistical evidence has failed to show that the disparity is not the result of random chance. And if random chance "caused the underrepresentation complained of,"⁵⁸ then the alleged defect in the jury-selection process did not. A defendant who has failed to demonstrate that observed underrepresentation is statistically significant (and offered no direct evidence of causation⁵⁹) has therefore failed to show, as required by Supreme Court precedent, that the underrepresentation is "systematic—that is, inherent in the particular jury-selection process utilized."⁶⁰ This reasoning suggests that a showing of statistical significance is sometimes necessary to satisfy *Duren's* "systematic exclusion" prong, even though such a showing does not establish whether *Duren's* substantial underrepresentation prong has been satisfied.⁶¹

V. CONCLUSION

Perhaps the most forceful objection against the disparity-of-risk test is that it complicates an area of constitutional law already plagued by numbers and equations. Yet this problem is easily solved. Courts can approximate the disparity-of-risk test by combining the relatively simple and already widely employed absolute and comparative disparity tests.⁶² Alternatively, courts could use the table from Part III to approximate the disparity-of-risk test without any calculation at all.⁶³

In any event, some added complexity is a price well worth paying in exchange for accuracy. Justice Breyer gave voice to this sentiment in the *Berghuis v. Smith* oral argument. After noticing the confusion caused by his invocation of the binomial theorem,⁶⁴ Justice Breyer worried that he would "hate to write something, like, saying 2 and 2 is 6," at which point the Chief Justice reassuringly noted that, because *Smith* was an AEDPA case, "all you have to do is say 2 plus 2 is somewhere between 3 and 5."⁶⁵

⁵⁸ *Duren v. Missouri*, 439 U.S. 357, 369 (1979).

⁵⁹ For an example of case involving direct evidence, see *United States v. Jackman*, 46 F.3d 1240, 1242–43 (2d Cir. 1995) (venire underrepresentation caused by computer malfunction).

⁶⁰ *Duren*, 439 U.S. at 366.

⁶¹ *Id.* at 364. For an overview of *Duren's* three-prong test, see *supra* note 1.

⁶² See *supra* pp. 12–16.

⁶³ See *supra* Table 1 and accompanying text.

⁶⁴ See *supra* notes 9–11 and accompanying text.

⁶⁵ Transcript of Oral Argument, *supra* note 6, at 6–7.

The Court was ultimately able to decide *Smith* without resolving the statistical questions raised during the case's oral argument.⁶⁶ Unfortunately, this meant that *Smith* had no occasion to supply guidance in this difficult area, even as it recognized that all the approaches currently "employed or identified in lower federal court decisions" are "imperfect" and sometimes "misleading."⁶⁷ Moreover, *Smith* was clear that "neither *Duren* nor any other decision of this Court specifies the method or test courts must use to measure the representation of distinctive groups in jury pools."⁶⁸ The Court will eventually have to confront the uncertainty created by its case law. And when it does, a statistically sophisticated approach to the fair cross-section requirement may be in the cards.

⁶⁶ *Berghuis v. Smith*, 130 S. Ct. 1382, 1394–95 (2010).

⁶⁷ *Id.* at 1393.

⁶⁸ *Id.*