THE TROUBLE WITH STATISTICAL EVIDENCE

RICHARD SCHMALBECK[*]

I

INTRODUCTION

Statistical evidence—frequently known as mere statistical evidence—has suffered at the hands of its critics for some time now.[1] The preceding piece by Professor Judith Jarvis Thomson[2] continues to refine the attack on such evidence—at least where it is the only evidence to support a court's judgment. In candor, I must admit that I am far from sure that Professor Thomson's attack is wrong, though I think it probably is. Perhaps it would be better to say that I find her primary argument unconvincing.

Professor Thomson mounts her argument largely in three hypothetical cases to which she applies the usual rules of evidence (or, in the last case, epistemology) to reach the conclusions that a civil defendant should be found liable, that a criminal defendant should be found guilty, and that a fact may be said to be known. Professor Thomson regards these conclusions as highly unsatisfactory, however, and isolates the problem in the statistical nature of the evidence supporting each result. She offers the implication that statistical evidence, at least by itself, provides insufficient grounds for a verdict in favor of the proponent of that evidence, primarily because it lacks a causal link to that which is to be proved. What is needed, she argues, is "individualized evidence," which, by her definition of the term, provides the necessary causal linkage.

A very brief recap of the three central hypotheticals may be appropriate. Of the three, the first seems clearly the most important. It involves a hit and run automobile accident caused by a recklessly driven taxi. Identification of the taxi proves to be impossible, but it is known that the Red Cab Company owned and operated sixty percent of the taxis in circulation on the night in question. Since by assumption no other relevant facts are known, it is asserted that the probability that the accident was caused by Red Cab Company is sixty percent.[3] Because this evidence meets the preponderance

1. See, e.g., Tribe, Trial by Mathematics: Precision and Ritual in the Legal Process, 84 HARV. L. REV. 1329 (1971).

2. Thomson, Liability and Individualized Evidence, LAW & CONTEMP. PROBS., Summer 1986, at 199.

3. As I argue below, this conclusion may not be as well-grounded as it seems. See infra pp. 227-29.

of the evidence standard, it appears that liability may be imposed on Red Cab Company. Professor Thomson suggests that we properly feel uncomfortable with this outcome.

The second hypothetical explores the use of statistical evidence in a criminal case. Two individuals set out independently to murder a third. One of the two succeeds, but no one can establish definitely which of the two is the actual murderer. It is known, however, that one would-be murderer fired a shotgun filled with five pellets in the victim's direction, while another, at about the same time, fired a shotgun filled with ninety-five pellets. On these hypothetical facts, it is asserted that the probability that the latter gunman is the murderer is ninety-five percent.[4] If a ninety-five-percent probability is taken to be an acceptable approximation of proof beyond a reasonable doubt, it would appear that the individual who shot ninety-five pellets could be convicted of murder. Again, we are told that this is not an acceptable result.

Finally, we are asked to consider a possibly analogous problem faced by a student of the theory of knowledge. Under the traditional view, in order to say that one knows that something is true, three conditions must be met: (1) it must actually be true; (2) the person averring knowledge must believe that it is true; and (3) that person must have a good reason for believing that it is true. Under these conditions, may A say that he knows B will lose a game of chance if: (1) B actually will lose; (2) A believes that B will lose; and (3) A's belief is based on the very substantial odds (perhaps nineteen to one) against B's winning? Arguably, these facts satisfy the specified conditions, yet it seems clearly wrong for A to assert under these circumstances that he *knows* B will lose.

It is, of course, not difficult to quibble with hypothetical cases. They are only metaphors for the real world, not the real world itself. It will always be possible to find ways in which any hypothetical case misrepresents a generally accepted reality. So I hesitate to embark on a dissection of Professor Thomson's hypotheticals. Unfortunately, I see no alternative. She structures her argument around the notion that the results in her hypotheticals make one uncomfortable, and she implies from those reservations that the shortcomings of statistical evidence are responsible for the unsatisfactory aspects of the results.

My objections to Professor Thomson's arguments partly stem from the fact that the outcomes of her hypotheticals disturb me much less than they appear to disturb her. For the most part, however, I base my objections on the view that whatever might be disturbing in her hypothetical cases is not derived from the statistical nature of the evidence in those cases, but rather from other defects in the particular bits of evidence, and in the interpretation of those bits.

I hope to make my objections clear largely through a detailed examination of the first hypothetical case, which involves Red Cab Company. References

---

4. Again, this conclusion may be suspect. *See infra* p. 231.

to the second and third cases will be made as they appear useful, but I will treat those cases less systematically.

II

IS IMPOSING LIABILITY ON RED CAB AN UNACCEPTABLE OUTCOME?

While it is true that the evidence against Red Cab is not wholly satisfying, it is far less clear that the outcome itself is wrong. If one accepts the level of the evidence as given, the only alternative outcome is that the plaintiff must bear the full brunt of the loss she suffered, even though that loss certainly wasn't her own fault, and is, we are told, more likely than not the fault of Red Cab under the respondeat superior doctrine.[5] The choice is between inflicting a possible injustice on Red Cab, and committing the injustice of leaving the innocent accident victim uncompensated. The magnitude of the injustice is about the same in either case.[6] If one focuses only on the outcome of the case, one must surely conclude that a finding of liability produces a better result than a finding of no liability.

Of course, the secondary effects of this outcome must also be considered. There are at least two ways in which the outcome may send the wrong signals to parties who may face identical or analogous situations in the future. First, one is troubled by a nagging sense that the liability imposed in these cases may be systematically excessive. If Red Cab Company is liable here, it presumably would be liable for all similar hit-and-run accidents involving unidentified taxis. This would be the wrong outcome, because the meager evidence we have suggests at most that Red Cab Company may have caused sixty percent of such accidents, but it will end up bearing one-hundred percent of the costs of those accidents.

To state the problem in this narrow way seems largely to resolve it. The circumstances are so improbable that concern about multiple iterations of this factual pattern seems absurd. But before dismissing the argument on the grounds of the extreme improbability of its facts, one should consider whether variants of Professor Thomson's hypothetical might be somewhat more probable. In fact, at least one variant has actually produced some amount of litigation—the case involving an allegedly defective product whose

---

5. The doctrine of respondeat superior is an application in the employment context of the general rule that a master is liable for injuries to third persons caused by the acts or omissions of his servants. RESTATEMENT (SECOND) OF AGENCY §§ 219-20 (1958). The employment relationship presumed in the Red Cab hypothetical establishes Red Cab's control over the job-related activities of its employees. Assuming that the hit-and-run injury arose from an activity in the course and scope of a driver's employment, the company would be liable for the injury. *See, e.g.,* Evington v. Forbes, 742 F.2d 834, 836 (4th Cir. 1984).

6. During the symposium at which this paper was presented, Professor Thomson suggested that there might be some additional harm—beyond the dollars lost in any judgment—to the reputation of the Red Cab Company. I doubt, however, that the harm would be significant in this case, since liability for automobile accident losses caused by its drivers must be a fairly routine event in the operation of a taxi company. More generally, one can imagine losses to reputation as a by-product of litigation for both defendants and plaintiffs; but this provides no basis for systematically weighting the outcome possibilities in either direction.

particular manufacturer cannot be certainly identified, though it is known that the manufacturer must be one of a discrete number of firms in the business of making the particular product at the relevant time. If Johns-Manville made sixty percent of the asbestos insulation material sold during some period, or Eli Lilly made sixty percent of the DES tablets prescribed, should they be liable for *all* of the damages caused by asbestos insulation and DES, respectively?[7] Cases raising this question have actually been litigated, and at least one potential solution—allocating liability pro rata according to market share—seems more satisfactory than making the predominant defendant bear the entire load.[8]

Allocating liability is a serious problem, surely worthy of detailed consideration.[9] It does not, however, seem to be what troubles Professor Thomson in her Red Cab Company hypothetical. She is not principally concerned with apportionment of damages; rather, she objects to the use of purely statistical evidence to establish *any* liability. She is untroubled by damage allocation questions because they don't get to the heart of her statistical evidence problem; I am untroubled because the emerging *Sindell*[10]—type approach to this problem seems reasonably acceptable. It will be better, then, to move on to the things that trouble at least one of us.

Another possible concern, also related to how Red Cab liability would affect incentives, has to do with litigation strategies. Even those who prefer to find the defendant cab company liable in this hypothetical must admit that the evidence is very thin. In fact, it seems unnecessarily thin. The plaintiff probably could have bolstered her case by, for example, examining the driving records of Red Cab Company's drivers, especially those on duty that night, or by examining the logs of trips taken that night. A rule permitting recovery in the Red Cab case might encourage less than optimal development of evidence by plaintiffs as a group. Certainly, the converse is true: dismissing this case because of the anorexic state of the evidence would encourage a heartier development of evidence by future plaintiffs.

But the argument is one-sided because it looks only at the plaintiff's incentives. The defendant in the Red Cab Company case also seems to have some opportunities to develop evidence in this case, and dismissal under the assumed state of the evidence may encourage less than optimal development of defendants' cases. The plaintiff, however, does have the burden of proof. If neither side presents any significant evidence, the plaintiff will lose. If, on the other hand, this plaintiff persuades the fact finder that it is more probable than not that the defendant was responsible for her loss, it is difficult to see any reason to require her to do more. Encouraging plaintiffs to do more permits defendants to do less, without any justification for such non-

---

7. These percentages are purely hypothetical.

8. *See, e.g.,* Sindell v. Abbott Laboratories, 26 Cal. 3d 588, 607 P.2d 924, 163 Cal. Rptr. 132 (1980).

9. *See* Robinson, *Multiple Causation in Tort Law: Reflections on the DES Cases,* 68 VA. L. REV. 713 (1982) (a detailed discussion of problems concerning liability allocation and multiple tortfeasors).

10. Sindell v. Abbott Laboratories, 26 Cal. 3rd 588, 607 P.2d 924, 163 Cal. Rptr. 132 (1980).

neutrality.[11] And, again, it does not appear that this is a problem about which Professor Thomson is particularly concerned. The structure of her hypothetical suggests that she hoped to present a case in which there simply was not much evidence that could be produced, not one in which the parties were indolent.

Based on my disposition of the possible secondary effects of imposing liability on Red Cab Company, I feel reasonably comfortable in asserting that the outcome of this hypothetical is more satisfactory than the alternatives in terms of the results to the immediate parties. In addition, it poses no important or unresolvable problems in terms of incentives to similarly situated parties in future cases.

III

IF THE OUTCOME IN RED CAB IS PREFERABLE TO ANY OBVIOUS ALTERNATIVES, WHY IS IT NEVERTHELESS UNSETTLING?

Although imposing liability on Red Cab seems preferable to the alternatives, I must admit that there is something troubling about the evidence. Professor Thomson argues that the evidence is troubling because it is statistical and therefore lacks the kind of causal connection with the accident that could be provided by individualized evidence. Perhaps that is the source of the problem, but I see other defects in the particular evidence presented in the Red Cab hypothetical that seem to me to offer better explanations of why the case is troubling. There seem to be three distinct types of problems—or perhaps three distinct ways of describing a single problem—with the statistical evidence in this case.[12]

1. *Confidence Level.*   The Red Cab case is presented as one in which the evidence that sixty percent of the taxis on the street on a particular night were Red Cab taxis supports the conclusion that, if nothing else is known, our best estimate of the probability that any taxi-caused damage that night was caused by a Red Cab taxi is sixty percent. The problem with this logic is that our best estimate here is not a very good estimate. The reason that it is not a very good estimate does not have anything to do with the fact that a sixty-percent probability is much less than a one-hundred-percent certainty. The problem is that one is not very confident that the estimate of sixty percent is correct.

It may be useful to distinguish two tiers of probability applicable here. The first is the probability of the event; the second is the probability that the first probability has been estimated accurately within some acceptable bounds.

---

11. The burden of proof itself operates to discourage frivolous lawsuits. But if the plaintiff can sustain that burden, the lawsuit is evidently not frivolous, and should not be discouraged.

12. The three "problems" I describe below have much in common, and it may be more accurate to think of them as a single problem viewed from three different vantage points. Even so, since some of these vantage points produce arguments that are more persuasive than others to some readers, it is worthwhile to separate them for discussion.

A concrete example of the two tiers of probability can be provided by an illustration from the statistical subfield that analyzes sampling error.

Imagine the case of a quality control inspector employed by a company that has just purchased 10,000 light bulbs. The contract provides that the 10,000-bulb shipment can be refused if more than four percent of the bulbs are defective. The inspector opens the first carton, and inspects fifty bulbs. He finds three defective. May he refuse the shipment? At that point, his best estimate is that six percent of the bulbs are defective. But he should feel uneasy about rejecting the shipment on that basis. His best estimate at that point is not a very good estimate.

He can do two things to improve the confidence in his estimate: he can inspect more bulbs, or he can inspect randomly selected bulbs, rather than ones chosen from a single carton. If he inspects one thousand randomly inspected bulbs and finds sixty defective, he is on much sounder ground. Even though his estimate of the probability that a given bulb will be defective has not changed, his confidence in that estimate has been greatly increased. Graphically, the difference in the two samplings described could be represented by a curve showing the true defect rate probability space that can be inferred from each:[13]

TABLE 1

LIGHT BULB DEFECT RATES

| Small Sample | | | Large Sample | | |

| Defect Rate | 6−X% | 6% | 6+X% | 6−X% | 6% | 6+X% |
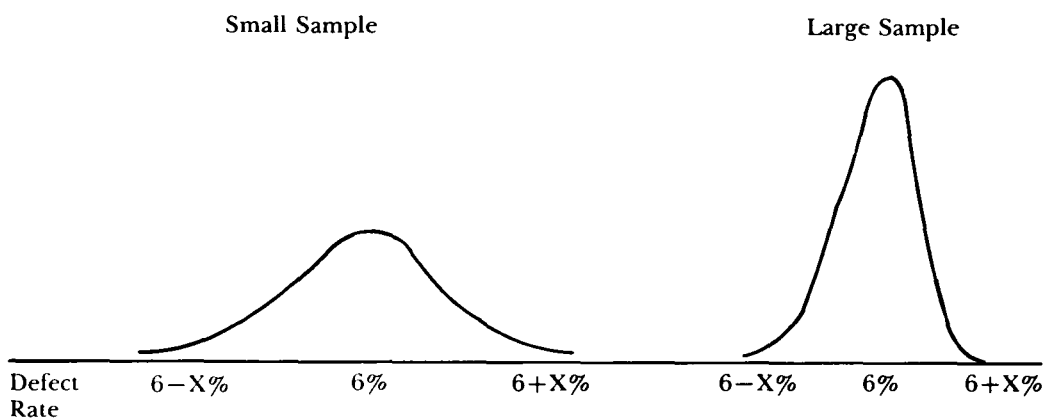
In both samples, the probability space centers on a defect rate of six percent; but in the larger, random sample, the inspector can say much more confidently that the defect rate is greater than the contractual maximum of four percent.

In the Red Cab case, sampling itself is not a problem: the full population of relevant cabs is known. The case presents an analogous problem, however.

---

13. These graphs are not plotted with mathematical accuracy. The general idea—that a small sample produces a wider range of outcomes on either side of the mean than a large sample—is clear enough.

There is substantial reason for concern that the probability space spreads widely around the sixty percent best estimate. A judge or juror would not, in other words, have reason to feel very confident that the true probability of Red Cab responsibility in this case was really anything very close to sixty percent. Just as in the light bulb sampling example, there are some things that could narrow the inferential probability space. Accident rates for Red Cab and its competitor and the number of driving violations accumulated by each company's drivers would be relevant. Traffic patterns also would be relevant. For example, we may want to know whether one company primarily services downtown-to-airport runs, while the other primarily serves the intra-city point-to-point traffic.

It is quite possible that all of these additional bits of statistical evidence would not yield anything of any interest individually, in the sense that they might not alter the probability estimate that it was a Red Cab taxi that caused the accident. We might find, in other words, that Red Cab drivers have sixty percent of all taxi accidents, sixty percent of all cab driver tickets, sixty percent of all trips in this neighborhood, and so on. But by eliminating possible explanations of why Red Cab might have sixty percent of the cabs and yet not be sixty percent likely to have caused the accident, such evidence would increase a fact finder's confidence in finding Red Cab responsible. The increase in confidence is not measurable, as it is in the sampling case, but it would be palpable to a judge or juror.

The problem with Professor Thomson's hypothetical appears in a starker light at the opposite extreme: a case in which the level of confidence in a probability estimate is absurdly low. Suppose the plaintiff's lawyer had not even bothered to find out how many cabs were operated by each company, and simply invited the judge to look out the courthouse window and identify the first five cabs he saw. Suppose further that three of the first five are Red Cab Company cabs. With no other evidence available, the best estimate would be that sixty percent of the city's cabs were operated by Red Cab Company. And once that conclusion was reached, it would be more probable than not that an accident caused by an unidentified taxi was caused by a Red Cab taxi. But surely that conclusion would be too flimsy to be tolerated.

The continuum I have described thus extends from a case where hardly any statistical evidence is offered (but where the available evidence points to a sixty-percent probability of fault), to a case in which a substantial body of statistical evidence is introduced (also pointing to a sixty-percent probability of fault). The difference at the extremes in the acceptability of the evidence is largely a matter of the confidence that a fact finder may reasonably have that the probability of fault really is around sixty percent or, to be more precise, really is greater than the fifty-percent tipping point under a preponderance of the evidence test.

Unlike the light bulb sampling example, the confidence levels in the Red Cab case are not readily quantifiable, and every fact finder must make a decision about whether the preponderence confidence level has been

achieved—a decision that is ordinarily purely intuitive. Professor Thomson's Red Cab hypothetical involves evidence that provides a confidence level between the two extremes I have outlined. For reasons outlined in section II above, my judgment would be that the evidence is barely sufficient; but I could surely see how people might differ in judging the weight of the evidence. My point here is that if a fact finder believes that the evidence in Professor Thomson's hypothetical is inadequate, it is likely to be because of the low confidence level in that evidence, not because it is statistical rather than individualized evidence.

2. *Mistaking a Part for the Whole.*   Some evidence—including some statistical evidence—suffers from another problem that may be part of what troubles Professor Thomson about the Red Cab hypothetical. Evidence in a case may present a fairly firm probability as to one element in a string of events in a situation where there is no evidence as to the probabilities of the other events that may be crucial to the ultimate conclusion. It is tempting to say that "all other things being equal," or "absent further information," one can apply that partial probability to yield an estimate for the whole. But this is usually a mistake. There is some flavor of this mistake in the Red Cab case. Certainly the census of taxis is relevant to the attempt to use probabilities to identify an otherwise unidentified cab. But it is only one factor in assessing the probability of the ultimate question: whose cab caused the damage?

An example may illustrate the statistical defect here. Suppose two baseball teams are about to begin a three-game playoff series for the league championship. Two individuals wish to make a bet on the outcome of the series at odds that are set by experts. All they can find in their local paper, however, is the Las Vegas line on the first game of the series, which is that Team A has a sixty-percent chance of winning the first game. May the bettors extrapolate the first-game odds into a sixty-percent probability estimate for Team A in the series?

It is clearly unsound to do so. In fact, sixty percent turns out to be the wrong probability for the whole series under all natural assumptions about the probabilities for the other games. How the "all else being equal" proviso is interpreted is crucial. Does it mean that Team A also has a sixty-percent chance of winning the second game, and the third game (if necessary)? If so, and if the games may be treated as independent events, then the probability that Team A will win the series is about sixty-five percent, not sixty percent.[14] If "all else being equal" means that Team A has a fifty percent chance of winning in each of the remaining games, then the probability that Team A will

---

14. Team A can win the series by winning the first and second games, the first and third games, or the second and third games. The probability of winning the first and second games under the assumptions given is .6 x .6, or .36. The probability of winning the first and third games, while losing the second is .6 x .4 x .6 = .144. The probability of losing the first game, and then winning games two and three is .4 x .6 x .6 = .144. The sum of these possibilities is .648.

win the series is fifty-five percent, not sixty percent.[15] Other distorting factors can readily be imagined: Team A may use its best pitcher in game one, whereas Team B may be expected to use its best pitcher in game two. Alternatively, perhaps Team A will be at home for game one, but on the road for games two and three. These possibilities suggest that the second and third game odds would differ from either fifty percent or sixty percent and that the series odds would differ from the two earlier extrapolations as well. What this exercise indicates is that, by itself, a probability estimate as to part of an inferential string doesn't really yield much information about the probabilities relating to the whole string.

To be fair, I don't think that this statistical fallacy is at the center of what is troubling about the Red Cab case. It is at the edges, however. Suppose that evidence is added to Professor Thomson's hypothetical as follows:

> (1) The accident occurred at about 8:30 p.m. (2) Red Cab changes shifts between 8:00 and 9:00 p.m., resulting in a cab utilization rate of only fifty percent during that hour (because of refueling and driver-change down time). (3) Red Cab's competitor changes shifts between 10:00 and 11:00 p.m., and so is able to maintain an eighty percent cab utilization rate between 8:00 and 9:00 p.m.

Under these facts, only about forty-eight percent of the cabs on the streets of the city at the relevant time were operated by Red Cab.[16] Thus, the total census of cabs provided only one fact of a pair (the other being the time-sensitive utilization rate) needed to draw the correct inference, and that one fact is not very helpful by itself. Substituting a probability of part of the logical chain for the whole simply won't do.

Is statistical evidence peculiarly subject to the problem of part-to-whole extrapolation? Is it especially likely to present itself as apparently precise data that turns out to be a probability of a part masquerading as a probability estimate for the whole? Perhaps, but Professor Thomson hasn't given us any explanation of why this might be so. I am inclined to think the difficulty is better explained by contrasting circumstantial evidence with direct evidence, rather than by contrasting statistical evidence with individualized evidence. Even if one looks exclusively at eyewitness accounts, for example, the statement "I saw the defendant rob the bank" is a statement bearing directly on the ultimate issue in question, while the statement "I saw the defendant leaving the bank just after a robbery alarm sounded" is clearly part of a chain of evidence from which a case against a defendant could be constructed. But is there anything inherent in statistical evidence that makes this problem especially difficult? If so, it is not obvious to me.

---

15. This time, the probability of winning the first two games is .6 x .5, or .30. The probability of winning the first and the third is .6 x .5 x .5, or .15. The probability of winning the second and third games after losing the first is .4 x .5 x .5, or .10. The sum of the probabilities is exactly .55.

16. Suppose there are one-thousand cabs licensed in the city: 600 Red Cab taxis and 400 others. Red Cab's utilization rate of 50% during the relevant hour means that 300 Red Cab Company cabs are on the streets during that hour, while the competitor's utilization rate of 80% means that 320 of its cabs are on the streets at that time. Three-hundred divided by 620 is .484.

3. *Mistrust of Statistical Evidence.*   A third source of concern about the use of statistical evidence in cases like Red Cab, which in some sense follows from the discussion of the first two, is that such evidence is often treacherous. It presents the same problem as expert testimony does: conclusions are presented, and typically poorly explained, to fact finders who must process the information without the kind of intuitive churning with which other testimony can be analyzed. Analytic mistakes are readily made, even by experts, but are not readily detected by the generally innumerate laymen who decide facts in trials. A vivid example was presented recently by the valiant but ultimately hilarious attempt by a *New York Times* reporter to explain—with expert help—what the odds were that a single person would twice win seven-digit lottery prizes, as a New Jersey woman recently did.[17] Statistical evidence often seems fragile. Professor Thomson's hypothetical cases are very much of this type. If the shift-change explanation I suggested in the preceding section were accurate, for example, the sixty-percent probability estimate of the cab in question being a Red Cab collapses totally. If assumptions of the independence of linked events proves faulty, statistical evidence based on such assumptions becomes invalid.[18]

---

17.   The full text of this news item follows:

On second thought, New Jersey lottery officials say, chances of winning the millionaire drawing a second time are no worse than the 5.2 million to 1 faced by anyone who holds a ticket in the drawing.

When Evelyn M. Adams of Point Pleasant won for the second time, on Feb. 10, the lottery said that the chances against a double victory were a stupendous 17 trillion to one. Now, they say, that figure applies solely to a very narrow set of circumstances.

In fact, lottery officials and mathematicians said, although the chances against a one-time winner are no worse than the 5.2 million to 1, depending on how frequently each winner continues to bet, the odds could improve significantly.

"The odds against you don't get worse because you won before," the assistant chairman of the mathematics department at City College, Walter Daum, said, "The lottery has no memory."

Mrs. Adams won $3.9 million in the Pick-6 Lotto last October and, with her fiance, won a $1.5 million jackpot Feb. 10.

The deputy director for planning and research of the lottery division, Barbara Steele, said that when officials presented Mrs. Adams to the press Feb. 13, they tried to explain that the 17 trillion referred solely to the "narrow and pristine conditions" of winning twice by buying just one ticket in each of two separate drawings.

"Mrs. Adams was never in that position," she said, because "she's not your typical player." Mrs. Adams was betting $25 a week when she first won, then improved her odds by wagering $100 in each drawing.

For his part, an economist for the State Economic Policy Council, Dr. Jong Keun You, who was the source for the 17 trillion figure, said the odds on having a repeat winner were influenced by such factors as how many first-time winners there were, how much each bet each week and how long they all continued to play. The lottery officials said they had not surveyed the 132 first-time winners to find out the answers to those questions.

Dr. John Koza, chairman of Scientific Games of Norcross, Ga., which runs the lotteries in New Jersey and other states, said that as the number of drawings and winners continued to grow, the odds favored having someone hit it big twice.

"Sooner or later," he said, "it was going to happen to somebody."
N.Y. Times, Feb. 27, 1986, at B9, col. 2.

18.   Another case Professor Thomson cites, People v. Collins, 68 Cal. 2d 319, 438 P.2d 33, 66 Cal. Rptr. 497 (1968), involved the failure of the independence assumption. For a full discussion of this case, see D. STRIPINIS, PROBABILITY THEORY AND CIRCUMSTANTIAL EVIDENCE, 5-15 (1983).

Professor Thomson's second hypothetical presents, quite inadvertantly, a nice example of the difficulty of applying statistical information to actual cases. The example involved a murder committed by one of two assailants, where one fired a shotgun loaded with ninety-five pellets at the victim while the other fired a shotgun loaded with only five. I don't mean to quibble with this example for the purpose for which it was used: Professor Thomson means to ask whether a defendant in a criminal trial can be convicted on a purely statistical case for his guilt where the statistical case points to a ninety-five-percent probability. Surely there is some hypothetical that cleanly presents such an issue, which is all Professor Thomson really needs. I have doubts about her particular case, however. I do not know much about shotguns, but I believe that the pellets from each shot are clustered within a fairly well-bounded area, or "spread." In this case, one of those clusters would contain ninety-five pellets, and be relatively densely clustered. The other would contain only five pellets, and be sparsely clustered. If we know that one *and only one* pellet hit our victim—and not in an ear lobe or a fingertip but in some vital spot—one might well suspect that that pellet was part of a sparsely populated pellet cluster; that is, that it was part of the five-pellet shot, not the ninety-five pellet shot, which would probably have created multiple wounds.

The point here is a simple one: it is rarely easy to translate statistical information into a reliable statement regarding the probability of the ultimate question in a dispute. So if one is uneasy about convicting the defendant who shot ninety-five pellets in this hypothetical, it may be because of uncertainties—perhaps at an intuitive level—about whether the pellet count really translates to a numerically equivalent probability that this defendant's shot caused the victim's death. Whether or not that is true as to Professor Thomson's hypothetical of the shotgun pellets, it clearly is true that nonstatistician fact finders are skeptical about the usefulness of statistical evidence. I would suggest that the skepticism is entirely justified by the slippery and insidious condition of the path from raw data to refined statements of probability.

IV

A BRIEF LOOK AT "INDIVIDUALIZED" EVIDENCE

In one sense, I may seem to be making Professor Thomson's case in the foregoing: she has reservations about statistical evidence, and so do I. In fact, however, the points I am making are quite different from hers. I think that statistical evidence must be used very carefully, since it will frequently convey less information about the ultimate factual issues than it seems to convey. But if it is used carefully, I think that it is as useful and reliable as nonstatistical evidence.

Professor Thomson, on the other hand, argues that statistical evidence alone is not enough to support verdicts for plaintiffs or prosecutors. To carry their respective burdens, she asserts, moving parties should be obligated to

provide "individualized evidence." To understand why individualized evidence is so potent, one should first ask what exactly it is. But do not ask me, because I do not know. Professor Thomson explains the concept, but in a way that I find dimly illuminating, at best.[19] She explains that what she is looking for is an "appropriate causal connection,"[20] and continues as follows:

> What sort of causal connection would be appropriate? Well, if a witness came forward to say he saw the accident, and that the cab which caused the accident looked red to him, *then* we would have what would be called individualized evidence against Red Cab; and my suggestion is that that is because the accident-causing cab's actually being red (and therefore being Red Cab's) would causally explain its looking red to that witness.[21]

But this explanation is not very helpful. After all, could I not say in defense of the cab census evidence that it was causally connected, as follows:

> The plaintiff (who is also a witness) says that she saw the accident and that the automobile that caused the accident looked like a cab to her. This would be individualized evidence because the accident-causing car's actually being a cab (and therefore being sixty percent likely to be a Red Cab) would causally explain its looking like a cab to that witness.

I have made one substitution in Professor Thomson's formula: I have slipped in a probabilistic statement where one did not appear before. How important is that substitution? Professor Thomson presumably thinks that this is very important, but she has not said why. The reason that I believe that my probabilistic substitution is an acceptable move is that there are already a number of implicit probabilities built into Professor Thomson's example of individualized evidence.

Her causal chain depends primarily on the truth of the witnesses' testimony. But witnesses lie, make mistakes, are subject to bias, and so on. Surely some chance exists that one of those things could be going on in her individualized evidence example. Further, Professor Thomson admits that even in her example, not all red cabs need belong to Red Cab Company. A given red taxi could, perhaps, be a cab from another city doing a long-distance run, or as Professor Thomson suggests (somewhat less plausibly), something disguised as a red cab.[22]

Is there a fundamental difference between Professor Thomson's formulation and my own? I do not think so. Professor Thomson's causal link proceeds from the observation that the accident was caused by a red cab to the conclusion, let us say, that it was ninety-nine percent likely that the cab was operated by Red Cab Company (as in her example of *individualized* evidence). My alternative causal link (in her example of *statistical* evidence) proceeds from the observation that the accident was caused by a cab to the conclusion that it was sixty percent likely that the cab was operated by Red Cab Company. The two analyses are fundamentally the same.

---

19. Thomson, *supra* note 2, at 202-06.
20. *Id.* at 203.
21. *Id.*
22. *Id.* at 204.

Professor Thomson's next example of individualized evidence even more obviously involves a probability statement. She writes:

> Or if it turned out that Red Cab had given a party for its drivers on the evening of the accident, a party that turned into a drunken brawl, then too we would have what would be called individualized evidence against Red Cab . . . because the party would causally explain its having been a Red Cab which caused the accident.[23]

But what exactly is the signficance of this drunken brawl? Recall that the legal theory of this case is *not* based on direct negligence of Red Cab, (such as might be shown by their hosting a drunken brawl). Rather, the implicit assumption on which this hypothetical is based is that Red Cab will be held liable, if at all, under a respondeat superior claim based on the negligence of the unidentified driver. The significance of the drunken brawl (and I agree that it has significance) is simply that it increases the probability (compared to what one knows without this fact) that any particular erratic cab driver that night was an employee of Red Cab.

Could we not then say:

> If it turned out that Red Cab operated sixty percent of the cabs on the street that night, then we would have individualized evidence against Red Cab, because Red Cab's operating most of the taxis would causally explain its having been a Red Cab that caused the accident.

Here, the significance of the cab census is simply that it increases the probability (compared to what one knows without this fact) that any particular cab driver on the streets that night was an employee of Red Cab.

To illustrate the equivalent value of these two pieces of evidence (the drunken brawl and the cab census), imagine a slight variation of the drunken brawl situation. Suppose it is known that some relatively fixed proportion of all cab drivers on duty for either company are drunk at any given time. Suppose also that it was Red Cab's *competitor* that had the party that degenerated into a drunken brawl. We then know that Red Cab had sixty percent of all drivers on the road that night, with the usual proportion of them being drunk. We also know that its competitor had fewer cabs on the road, but an abnormally high proportion of them were driven by drunk drivers. Do we accord higher status somehow to the evidence against the competitor because it is individualized evidence under Professor Thomson's definition, whereas the evidence against Red Cab is merely statistical? I think not. More sensibly what we want to know in this case is how many drunk drivers each company had on the road that night. To make that determination, each piece of evidence (the competitor's party, and Red Cab's larger market share) is equally relevant.

By this point, I hope to have convinced at least some readers that the examples around which Professor Thompson has organized her paper are only superficially troubling. It is of course possible that other examples involving the use of statistical evidence might be more convincing, or at least better able to withstand the criticisms I have offered of her chief examples.

---

23. *Id.* at 203.

Although I am skeptical of the alleged inferiority of purely statistical evidence and indeed somewhat confused even about what should properly be considered purely statistical evidence, I have made a good faith effort to come up with cases in which I would be uncomfortable with proceedings based on simple probabilistic statements about a defendant's liability or guilt. The following hypothetical is the most troubling case which I could develop: Suppose the Internal Revenue Service has reliable aggregate taxpayer statistics indicating that sixty percent of taxpayers who have adjusted gross incomes between $100,000 and $200,000 overstate their charitable contributions deductions by at least $500. May the IRS properly assess a deficiency equal to the tax on $500 of additional taxable income with respect to any taxpayer within this income bracket?

By assumption, there is a sixty percent chance that at least that amount of tax would be owed; it is more likely than not, in other words, that any particular taxpayer within the relevant bracket should pay at least the additional tax assessed. The worrisome aspect of making such assessments, of course, relates to those taxpayers who fastidiously reported exactly the amount of their true charitable contributions and who may number up to forty percent of the relevant population. Is it appropriate to subject them—or, more accurately in any particular case, to *risk* subjecting them—to deficiency assessments on purely probabilistic evidence?

Even though I was initially troubled by this example, I ultimately concluded that, while a variety of administrative concerns in the present system militate against using this particular enforcement strategy, the assessment procedure described should not be viewed as impermissible. It must be kept in mind that the honest taxpayer has an opportunity to mount a defense based on his records, his own testimony, and the testimony of others. Such evidence should, and almost certainly would, outweigh statistical evidence of the sort described, which says no more than that some taxpayers more or less like the present taxpayer understate their tax liabilities. It may seem unfortunate that a taxpayer who has done nothing wrong should be put to the trouble of proving his innocence. However, our legal system generally, particularly the tax system, does not reflect great concern about defendants being put to the trouble of defending themselves. Any plaintiff can bring an action against any defendant. If the action is completely groundless, the defendant may be in a position to recover under a counterclaim based on abuse of process. Normally, however, the defendant simply wins by not losing, which means that he has suffered some damage in the form of his legal costs and his trouble. But that is accepted as one of the costs of giving plaintiffs liberal access to remedies.

The situation of a taxpayer with respect to the Internal Revenue Service is in some sense worse than the situation of defendants in general. In auditing taxpayers, which we may think of as constituting the prelitigation phase of tax controversies, the IRS simply summons some two million taxpayers per year and puts them to the trouble of showing that they reported their income and

deductions in a manner that was factually accurate and legally correct. In fact, the taxpayer's situation in the audit, or prelitigation, phase is similar to my hypothetical case of assessing a deficiency on the basis of probabilities, because the IRS does indeed select returns, and particular items on those returns, for the attention of auditors on the basis of probability functions that yield data about the amount and likelihood of misreporting on the initial return. If the taxpayer produces evidence—testimonial, documentary, or both—that the auditor finds credible, the tax case proceeds no further. If a taxpayer were to refuse to cooperate with an auditor, the IRS would normally proceed to assess additional taxes based on a disallowance of *all* unsupported deductions, not simply on the basis of the statistically predicted overstatement of deductions, putting the burden on the taxpayer of proving any or all of the deductions claimed. It would be *less*, not more, objectionable from the taxpayer's view to proceed on the basis of statistical predictions rather than relying on shifting the burden of proof. Yet shifting the burden of proof is the accepted practice, while statistical prediction is the "troubling case" under analysis. The "troubling case"—which was about the worst statistical evidence abuse I could think of—certainly seems less troubling when viewed from this perspective.[24]

V

CONCLUSION

Ultimately, I think my most basic problem with Professor Thomson's argument is that I cannot, from her definition and examples, make her "individualized evidence" concept yield anything more than that such evidence is relevant evidence. But relevant evidence is simply evidence that affects a fact-finder's judgment about the probability of guilt or liability. Relevant evidence can be either statistical or nonstatistical. Statistical evidence can be troublesome in part because it often *appears* to translate readily into a conclusion about the probability of the ultimate issue. The burden of the first two-thirds of this comment was to show that the appearance is often deceptive. But that doesn't mean that statistical evidence

---

24. One of my colleagues believes that I have missed much of the point of Professor Thomson's article, though he was too kind to put it in quite those terms. He thinks that the deeper concern reflected in Professor Thomson's article pertains to the tension between group rights (and responsibilities) and individual rights (and responsibilities). Professor Thomson's complaint, in his view, has to do with the use of group-generated probabilities to prove the liability or guilt of individuals. Not surprisingly, my colleague was particularly disturbed by the IRS hypothetical case in this section.

My colleague may well be correct in identifying the sensibility that drives Professor Thomson's article; to that extent I have perhaps not directly joined issue with her. My response to this objection is contained in section II of my comment. It seems wrong to permit concerns for the individual rights of the defendant to overwhelm an equally significant concern we should feel for the rights of the plaintiff to obtain redress. To be fair to Professor Thomson, she does not say that the outcomes of her hypotheticals are wrong—only that they are troubling. If that is all she means to say, I would have to agree with her. But, by itself, that observation is surely trivial. Hard choices—ones in which there is substantial risk of harming an innocent individual whichever way one proceeds—are always troubling.

is somehow of a lower order than nonstatistical evidence. The latter evidence, eyewitness accounts, for example, has its own sets of problems. I have mentioned them,[25] but they are sufficiently obvious that I need not belabor this point.

Professor Thomson draws a number of what I believe are false distinctions between statistical evidence and what she calls individualized evidence. She raises the concern that if we find Red Cab Company liable on the ground that it had more cabs on the road than its competitor, it will be mere luck if a Red Cab did indeed cause the accident. But what if we have a witness who says he saw the accident, who thinks that the cab that caused it was red, but isn't sure because it was a dark and stormy night illuminated only by mercury vapor lights that somewhat distort color perceptions? Such a witness' statements do not typically come with probability estimates appended, but it is clear that there exists in this case a substantial likelihood that the witness has misidentified the cab. In the face of that uncertainty, can one plausibly claim that it will be more than "mere luck" if a Red Cab actually caused the accident?

Inevitably, fact finders are engaged, either explicitly or implicitly, in something of a statistical exercise: they seek an estimate in which they can be reasonably confident of the probability of the defendant's guilt, negligence, or other basis for liability. The essentially probabilistic nature of this endeavor makes me doubt the relevance of the last of Professor Thomson's major hypotheticals—the theory-of-knowledge hypothetical involving a person who claims to "know" that another will lose a lottery when there are inadequate grounds to support such knowledge. Rarely, if ever, are fact finders in a position to *know* whether a defendant is guilty or negligent. Wisely, the judicial system does not require knowledge. It requires only probabilities: it must be more likely than not that the defendant was negligent; or, it must be likely beyond a reasonable doubt that the defendant was guilty. Professor Thomson is quite right in thinking that statistical evidence of the sort exemplified by her hypotheticals does not yield knowledge. But if properly used it can yield a sense of the relevant probabilities, and that is all that can be asked of any evidence, statistical or otherwise.

---

25. *See supra* p. 231.