Is it a Crime to Belong to a Reference Class?*

MARK COLYVAN
Philosophy, University of Tasmania

HELEN M. REGAN and SCOTT FERSON
Applied Biomathematics, New York

I. INTRODUCTION

ON DECEMBER 10, 1991 Charles Shonubi, a Nigerian citizen but a resident of the USA, was arrested at John F. Kennedy International Airport for the importation of heroin into the United States.[1] Shonubi's *modus operandi* was "balloon swallowing." That is, heroin was mixed with another substance to form a paste and this paste was sealed in balloons which were then swallowed. The idea was that once the illegal substance was safely inside the USA, the smuggler would pass the balloons and recover the heroin. On the date of his arrest, Shonubi was found to have swallowed 103 balloons containing a total of 427.4 grams of heroin. There was little doubt about Shonubi's guilt. In fact, there was considerable evidence that he had made at least seven prior heroin-smuggling trips to the USA (although he was not tried for these). In October 1992 Shonubi was convicted in a United States District Court for possessing and importing heroin. Although the conviction was only for crimes associated with Shonubi's arrest date of December 10, 1991, the sentencing judge, Jack B. Weinstein, also made a finding that Shonubi had indeed made seven prior drug-smuggling trips to the USA.

The interesting part of this case was in the sentencing. According to the federal sentencing guidelines, the sentence in cases such as this should depend on the total quantity of heroin involved. This instruction was interpreted rather broadly

in the Shonubi case so that the total quantity of drugs should include all eight of Shonubi's drug-smuggling episodes. It was estimated that the total quantity of heroin that Shonubi carried into the USA on his eight trips was $427.4 \times 8 = 3{,}419.2$ grams. This was above the crucial 3,000 gram threshold and so corresponded to a base offence level of 34 (which, in Shonubi's case, resulted in a sentence of 12 years and 7 months imprisonment). (Call this *Shonubi I.*)

Shonubi appealed this sentence and the Second Circuit Court of Appeals vacated the sentence because the total quantity of drugs smuggled by Shonubi had not been established by a preponderance of evidence.[2] This was in part because there was no "specific evidence" that Shonubi smuggled the quantity of drugs on which the sentencing was based. The case was sent back to Judge Weinstein in the District Court for resentencing. (Call this *Shonubi II.*)

In response to the lack of specific evidence, the prosecution offered data collected by the US Customs Service. This data consisted of the quantities of heroin seized from 117 Nigerian balloon-swallowing drug smugglers arrested at JFK airport between the dates of Shonubi's first- and last-known drug-smuggling trips: September 1, 1990 and December 10, 1991. The prosecution statistician used repeated sampling (with replacement) from the 117 data points and produced a Gaussian-like histogram for the quantity of drugs imported on Shonubi's seven prior trips. The statistician concluded that there was a 0.99 chance that on the seven previous trips Shonubi smuggled at least 2,090.2 grams of heroin. When this was added to the 427.4 grams found on Shonubi on his last trip, the total quantity of drugs imported by Shonubi was estimated to be about 2,500 grams. Based on this evidence, Judge Weinstein found that Shonubi smuggled between 1,000 and 3,000 grams of heroin in his eight trips for a base offence level of 32 which (after an enhancement added for Shonubi's lies and obstruction of justice) resulted again in a sentence of 12 years and 7 months imprisonment. (Call this *Shonubi III.*)

Once again Shonubi appealed and once again the sentence was vacated by the Second Circuit. Two very interesting points were made by the appellate court at this stage. The first was to reiterate the requirement for "specific evidence" and to point out that the Customs Service data did not constitute such evidence. The second point was that preponderance of evidence was too low a standard of proof for sentencing matters relating to disputed aspects of relevant conduct. In particular, when a significantly enhanced sentence was at stake, a more rigorous standard of proof was necessary. The court then referred the case back to the District Court for resentencing with the additional instructions that the prosecution had had two previous opportunities to provide specific evidence of

---

[2]There are three different standards of proof appealed to (in decreasing order): beyond reasonable doubt, clear and convincing evidence, and preponderance of evidence. While beyond reasonable doubt is required for conviction, sentencing requires only preponderance of evidence.

the total quantity of drugs imported by Shonubi and no further opportunity to provide such evidence would be allowed. (Call this *Shonubi IV*.)

Back in the District Court, Judge Weinstein had no choice but to sentence Shonubi based only on the 427.4 grams found on him on his last trip. This resulted in a sentence of eight years and one month imprisonment. (Call this *Shonubi V*.) The trial then moved out of the court and into the academic arena where it continues to this day.

The general consensus amongst commentators on this case seems to be that the Second Circuit Court's decision at *Shonubi IV* to vacate the sentence of *Shonubi III* was a poor one. In particular, most commentators suggest that this poor decision was due to lack of understanding of statistical methods. Moreover, some suggest that the real importance of the Shonubi case is in the lessons it holds for statistical evidence generally. Here is what some of the commentators have said:

> The opinion of the court of appeals is interesting, important, and—above all— depressing because of what it suggests about the difficulties that many judges may have when they confront statistical methods and statistical evidence: it suggests that quite a few judges—including some very eminent and intelligent ones—still may not have a grasp of some basic characteristics of probabilistic and statistical methods and arguments.[3]

> [T]he opinion of the court of appeals in *Shonubi IV* is not a distinguished one . . . [T]he Second Circuit's view that the statistical evidence relied upon by Weinstein was defective because it was not "specific" is most unfortunate. The notion that evidence of a person's actions must be somehow "specific," either to that person or to his [sic] behavior on a specific occasion, is at best otiose and, at worst, nonsensical.[4]

> The Shonubi case illustrates the legal system's failure to fully appreciate statistical evidence.[5]

Judge Weinstein even suggested that the decision of *Shonubi IV* "represents a retrogressive step towards the practices relied upon from the Middle Ages to the late Nineteenth century."[6]

Now it seems that there is much more at issue here than the sentencing of Charles Shonubi. If we are to accept what some of the commentators are saying, the use of statistical evidence in US courts is on trial. The concern is that the rejection of the statistical evidence in this case amounts to the wholesale rejection of statistical evidence (presumably because the statistical evidence here is as good as statistical evidence ever gets). We disagree with the commentators above on two counts. First, we believe that the Second Circuit Court's decision was not a poor one—the reasons given for the decision were perhaps not expressed as well

---

[3]Tillers 1998.
[4]Ibid.
[5]Gastwirth, Freidlin and Miao 2000, p. 405.
[6]Quoted in Tillers (1998).

as they might have been, but the idea was right. Second, we do not believe that the statistical evidence in this case was as good as statistical evidence gets, so there is no serious threat to the future of statistical evidence in US jurisprudence.[7]

For the most part, we defend the Second Circuit's decision by criticizing the statistical evidence accepted at *Shonubi III*. Along the way, however, we say a little to defend the reasons given by the Second Circuit Court for rejecting this evidence. In the next section we discuss our main objection to the evidence presented at *Shonubi III*: the evidence founders on the well-known reference-class problem. In the following section we discuss the Shonubi case from a decision-theoretic angle and suggest that the close attention to the statistical evidence in this case has blinded many to the importance of the relevant utilities.

II. THE REFERENCE-CLASS PROBLEM

There were several serious flaws in the methods employed by the District Court at *Shonubi III* in determining the quantity of drugs on which Shonubi's sentencing was based.[8] Perhaps the most serious of these flaws was that no estimates of measurement error or other uncertainties about the quantities of drugs recovered were recorded in either Shonubi's case or for the other points in the data sets. Moreover, these errors were likely to be rather significant once the methods of determination are considered.[9] Moreover, the fact that these errors were significant should have been obvious, given that 17 of an original 142 data points from the Drug Enforcement Administration contained inconsistent information (net weights greater than or equal to gross weights).[10] We shall not dwell on such matters here. Although these details are important, the real issue is whether Shonubi should have been sentenced based on evidence gathered from *other people*. That is, should he have been sentenced based on his membership in a particular reference class?[11]

First we note that although Shonubi was not *convicted* because of his membership in a reference class, this possibility is not ruled out. The only

[7]See Tribe (1971) for a somewhat more pessimistic view of the proper role of mathematics in the legal process.

[8]Leaving aside the obviously flawed method of multiplication by 8 used in *Shonubi I*.

[9]See Izenman (2000a) for details of the various shortcomings in the estimate of the quantity of heroin found on Shonubi on December 10, 1991—the date for which he was convicted.

[10]Although these data were not used as evidence (it was the Customs Service data that were ultimately used as evidence), the Drug Enforcement Administration data were used in determining the relationship between gross and net weights in the Customs Service data.

[11]There are in fact two issues here: the choice of reference class and the specificity of the evidence. These issues can, of course, come apart. For example, consider a case where the reference class in question consists of patterns of an individual's past behavior. Evidence obtained from such a reference class, we take it, would pass for specific evidence, but would still suffer from concerns about the choice of reference class. Evidence obtained from the reference class in the Shonubi case, however, was clearly non-specific, since it arose from data collected from individuals other than Shonubi. For the most part, our discussion focuses on the reference-class issue although our concerns are motivated by the non-specific nature of the class used at *Shonubi III*. (We discuss the specific evidence issue in more detail toward the end of this section.)

difference between the standards of evidence required for conviction and sentencing is that the former employs the beyond-reasonable-doubt standard, while the latter employs the preponderance-of-evidence standard. So, for instance, if the probability that Shonubi carried more than 2,000 grams of heroin into the USA was sufficiently high (0.99, say) then presumably this would suffice for a conviction.[12] That is, if this kind of evidence is to be accepted, as Tillers and others suggest, we ought to accept it for conviction as well, in which case it *would* appear to be a crime to belong to a reference class. So, for example, let us suppose that 99 per cent of people from a certain reference class cheat on their taxes. Does this mean that we are justified in charging and sentencing someone in this class with tax evasion, without further evidence? No, of course not; we require more evidence than simply their membership in the reference class in question. It is important to note that we require further evidence *not* because we wish to raise the probability from 0.99 to something higher (after all a probability of 0.99 seems a good candidate for beyond reasonable doubt). Rather, we require further evidence because the reference-class evidence is not specific to the individual in question.

The problem here is rather well known and it is somewhat surprising that the critics of *Shonubi IV* do not acknowledge this.[13] The problem is, of course, the reference-class problem. To appreciate the problem consider the following defense of Shonubi's innocence. Shonubi is a toll collector at the George Washington bridge. Let us suppose that no other toll collectors at this bridge were engaged in drug smuggling. Now Shonubi could appeal to his membership in *this* reference class to protest his innocence (or to argue that he imported considerably less than 1,000 grams of heroin in his eight trips). The point is simply that Shonubi is a member of many (in fact infinitely many) reference classes; some of these classes consist largely of unsavory types while others consist largely of saints. Membership in reference classes, it seems, does little to establish anything about Shonubi's own behavior.[14]

We are not claiming that there is no solution to the reference-class problem, just that there is no straightforward solution, and until a satisfactory solution is provided, the difficulty we raise here is serious. As an illustration of how deep the reference-class problem is, we will discuss and put to rest a couple of possible candidates for solutions that might suggest themselves in the case at hand. The

---

[12]Of course the reference class used to derive the statistics at *Shonubi III* presupposed Shonubi's guilt and indeed his guilt seems beyond question. The point remains, however, that if we accept such statistical evidence and such evidence could be obtained *without* presupposing his guilt, that data would, it seems, suffice for a conviction.

[13]Actually Tillers (1998) does discuss this issue briefly but he does not pursue it because he does not believe that this is the reason for the Second Circuit Court's rejection of the customs service data.

[14]Note that it is not simply a question of getting the reference classes tight enough. That is, it is not just a question of specifying enough predicates to be jointly satisfied so that the reference class in question contains very few (but non-zero) members. The size of the reference class is irrelevant (unless it contains just Shonubi); we require homogeneity of the class and this, in general, has nothing to do with its size.

basic idea is that the reference class of Nigerian balloon-swallowing heroin smugglers is seen, by some, to be intuitively relevant, whereas the class of toll collectors, for instance, is not. We need to find an argument in support of this intuition, though.

The first suggestion for justifying the reference class of Nigerian balloon-swallowing heroin smugglers might be that it was Shonubi's membership in this class that landed him in court in the first place.[15] This, however, is simply not true. It was Shonubi's membership in the class of apprehended drug smugglers that landed him in court. The mode of transport, country of origin, and, indeed, the kind of illegal substance were not relevant—he would have ended up in court if he were caught smuggling cocaine from Australia in stuffed toys, say. Indeed, he would have ended up in court if he had been caught engaging in any illegal activity. We take it, however, that the reference class of apprehended criminals is not intuitively the appropriate one, so this justification of the reference class in question will not do.

Next consider the suggestion that the class of balloon-swallowing heroin smugglers is more robust than others. That is, taking the intersection of this class with others (such as the class of toll collectors) will not significantly change the estimate of the quantity of drugs smuggled. If correct, this amounts to the claim that the reference class in question is homogeneous with respect to the quantity of drugs smuggled. The problem is that such homogeneity has not been established by any of the statistical evidence presented, and to simply assume homogeneity is to beg the question. Whether this reference class is homogeneous, and, in particular, whether it delivers the correct answer for Shonubi, is precisely what is at issue. We cannot invoke such unsupported assumptions to justify the appropriateness of this reference class.

Finally, consider the suggestion that there is a causal story to be told here: there is a causal connection between Shonubi's membership in the reference class in question and his carrying the quantity of heroin with which the District Court sentenced him at *Shonubi III*. The story would, in fact, be a story about a common cause: perhaps all members of the reference class in question worked for the same drug lord, say, who ensured that all drug couriers carry, more or less, the same quantity of heroin on each trip. At the very least, you might believe there to be a counterfactual dependence between carrying a certain quantity of heroin and being a member of the reference class in question: had Shonubi not carried the quantity of drugs on which he was sentenced at *Shonubi III*, he would not have been a member of the reference class in question. The problem with these responses is that in the former variant, causal connections that have not been established are being assumed, and in the second variant, homogeneity of the reference class is again being illicitly invoked. It seems that this justification also fails.

---

[15]We thank an anonymous referee for this and the next suggestion.

We do not wish to turn this discussion into a survey of proposed solutions to the reference-class problem; we merely wish to note that (i) the problem is serious and (ii) the obvious "solutions" fail. Whether there are more sophisticated strategies that can be made to work, we leave as an interesting open question. It is worth noting, however, that no justifications, sophisticated or otherwise, were provided by critics of the Second Circuit Court's decision at *Shonubi IV*. The reference-class problem, for the most part, was simply ignored. We now turn to the question of whether the reference-class problem is a problem only for frequentist interpretations of probability.

What is at issue here is the probability that Shonubi carried between 1,000 and 3,000 grams of heroin. We are not interested in how many others (no matter how much they are like Shonubi) carried between 1,000 and 3,000 grams of heroin. That is, we are interested in a single-case probability, not a frequency relative to some reference class. Once put this way it sounds as though what is at issue is the notorious problem of the correct interpretation of probability. According to frequentists such as von Mises (1957), we cannot even make sense of single-case probabilities, for probabilities are simply ratios of frequencies of certain kinds. (This view should be contrasted with other interpretations such as the subjectivist interpretation or the propensity interpretation, where single-case probabilities are countenanced.[16]) Indeed, such a frequentist view seems to be lurking in the background of the criticisms of *Shonubi IV*.[17] It would take us too far afield to defend an interpretation of probability theory that countenances single-case probabilities. Suffice to say, however, that it is hard to see how any view that did not allow single-case probabilities could be of any use in decision theory (which is, after all, what we are talking about here—we have more to say on this in the next section). In any case, the critics of *Shonubi IV* should at least acknowledge that there is considerable debate over the interpretation of probability; they should not simply assume a naïve frequentist view without comment.[18]

There is another reason for not embarking on a defense of single-case probabilities here: single-case probabilities may not solve the problem at hand. It might be argued that the reference-class problem also has considerable bite on interpretations of probability that countenance single-case probability statements. After all, even subjectivist and propensity views of probability should be answerable to the frequencies. For example, subjective probability assignments should (at least roughly) coincide with relative frequencies, when such frequency data is available, otherwise huge epistemic problems loom for these

---

[16]See Resnik (1987) for a brief introduction to some of the different interpretations of probability.
[17]For instance, recall Tiller's second passage above where he suggests that the request by the Second Circuit Court for specific evidence was "nonsensical" and that the rejection of the statistical evidence in this case may amount to "a blanket condemnation of statistics and statistical methods" (Tillers 1998).
[18]See Hájek (1997), Jeffrey (1992) and Papineau (1995) for criticisms of frequency interpretations of probability. See Carnap (1945) for the suggestion that there are two distinct conceptions of probability.

interpretations of probability. The point is simply that while non-frequency interpretations of probability deny that probabilities should be *identified* with frequencies, they ought not deny that probabilities are at least in some sense answerable to frequencies.[19] But this is enough for the reference-class problem to be reinstated, because the frequencies to which the non-frequency interpretations must answer suffer from the reference-class problem. For example, let us suppose that an agent A assigns the subjective probability s to some event e occurring under circumstances c. Now this subjective assignment might need to be revised when new evidence comes to light. So, for instance, let us suppose that it is found that e occurs p per cent of the time in circumstances like c and that $s \neq p/100$. It seems reasonable to suppose that A should revise s to something closer to $p/100$ (perhaps by Bayes's theorem). But the reference class used to determine p can be questioned. Had we used a different reference class we might have found that e occurs q per cent of the time (where $q \neq p$) and A's revision of s would have been different. It thus seems that the reference-class problem arises whenever relative frequencies are used—it is not simply a problem for the frequency interpretations of probability.[20]

It is interesting to note that although the Second Circuit Court did not mention the reference-class problem in their rejection of the evidence presented at *Shonubi III*, nevertheless it seems that their concerns were closely related to the reference-class problem (at least if it is thought of as a problem for the frequency interpretation of probability). But if what we have suggested above about the reference-class problem being a problem for single-case probabilities as well (at least for those that are answerable to frequencies), it might seem that we are committed to the wholesale rejection of statistics from legal evidence. This, however, is not the case. We now sketch a way of making sense of the Second Circuit's request for specific evidence that does not involve the rejection of all statistical evidence.

First, we note that no matter how you cash out the phrase "specific evidence," there is an obvious candidate for such evidence in the Shonubi case: Shonubi's previous behavior. This evidence might include previous convictions, records of financial dealings and so on. Now such evidence may not have actually been available in the Shonubi case but it is certainly not "nonsensical" to ask for such evidence.[21] It is important to note that the reference-class problem arises in relation to such specific evidence as well but at least the events in any such reference class would all be events involving Shonubi; he would not be sentenced based on the behavior of others, but on his own previous behavior.

Second, it is useful to bear in mind the distinction between inductive reasoning (typically involving frequency data) and abductive reasoning or inference to the

---

[19]For the related issue of how degrees of belief should be constrained by objective chance, see Beebee and Papineau 1997; Lewis 1986; Reichenbach 1949; Kyburg 1974; and Mellor 1971.
[20]We owe this point to Phil Dowe (1992). See also Hájek (forthcoming) and Ayer (1972).
[21]As Tillers (1998) suggests it is.

best explanation (which may involve probabilities interpreted subjectively or as propensities). For example, it seems that inference to the best explanation was employed to arrive at the conclusion that Shonubi was involved in seven previous drug-smuggling trips. That is, the best explanation for his multiple visits to Nigeria (and his multiple passports that presumably try to disguise this fact) on his relatively low income as a toll collector is that he was smuggling drugs on each of these occasions. Whatever problems inference to the best explanation may face,[22] it does seem to satisfy the constraint of specific evidence and that is all we are interested in here. (For example, the above argument for the conclusion that Shonubi was involved in seven prior drug-smuggling trips does not involve evidence pertaining to anyone other than Shonubi.) Now if we accept inference to the best explanation, we have another kind of evidence that might have been employed in the Shonubi case to support the prosecution's claims about the total quantity of drugs imported by Shonubi. This evidence might consist of economic considerations such as the minimum quantity of heroin needed to cover the costs involved in the smuggling operation.[23]

To sum up this section. We have argued that, while the Second Circuit's reasons for rejecting the statistical evidence of *Shonubi III* might not have been expressed as clearly as one would like, their decision was correct. The statistical evidence presented at *Shonubi III* ignored the well-known reference-class problem—statistical data were presented that were based on a particular reference class as though doing so was uncontroversial. One cannot simply assume that the reference class used is privileged, especially when significant increases in prison penalty may result. Finally, we deny that rejecting the evidence of *Shonubi III* amounts to the wholesale rejection of statistical evidence in law and we provide two examples of statistical (broadly construed) evidence that should count as specific in this context.

III. A DECISION-THEORETIC PERSPECTIVE

Another line of attack on the District Court's decision at *Shonubi III* is that we are not simply interested in the probability that Shonubi carried more than 1,000 grams of heroin into the USA; we wish to decide on some course of action and so we must also consider the utilities associated with the outcome of the proposed actions. When faced with a decision for which both the relevant probabilities and utilities are available, decision theory counsels us to choose the action that maximizes expected utility.[24] To ignore the importance (or even the existence) of utilities in any decision-making exercise is tantamount to accepting that all

---

[22]For some of these, see: van Fraassen 1980; and Cartwright 1983.

[23]We should point out that economic evidence was also rejected as non-specific by the Second Circuit Court. It is not clear what this evidence consisted of, but if it was the kind of evidence we are suggesting here, we would, in this instance, disagree with the Second Circuit Court's decision.

[24]Of course we may choose to use some other decision-theory rule such as maximin or minimax regret. We discuss these shortly. For the moment let us stick with maximizing expected utility.

consequences have identical value—that we are indifferent to the outcomes of the proposed acts. To illustrate how utilities can make a difference to a decision, suppose we are asked to walk across a narrow beam placed a few inches above the ground. If we successfully cross the beam, we win a small sum of money; if we fall off, then we miss out on the cash. In this case the utilities are such that we would elect to cross the beam provided we have no specific objection to beam walking. Now suppose the beam is placed high across a ravine, below which are crocodile-infested waters. Even though the probability of successfully crossing the beam is identical to that of the previous scenario (our beam-walking skills have not altered) the utilities have changed to such an extent that we are no longer inclined to accept the challenge.

Of course, the importance of utilities in judicial decision-making is recognized routinely. The option of a plea bargain in many criminal cases is based on maximizing expected utilities. The benefits to the prosecution and the courts aside, defendants often opt to plead guilty to crimes in order to receive a reduced sentence. In such cases, even though the probability of being found guilty has increased to (or, in some cases, remains at) a value of one, the expected utility, measured (inversely) in terms of jail time, is maximized with the act of pleading guilty. It is surprising, then, that criticism of the appellate court's decision in *Shonubi IV* was leveled squarely at issues relating to probabilities, with no regard to the relevant utilities. Tillers (1997; 1998), Izenman (2000; 2000a) and Girstwirth et al. (2000) all quite rightly emphasize the importance of probabilities in judicial decision-making, but they completely overlook the role of utilities. In decisions under risk, the utilities are as important as the probabilities. In fact, in decisions under ignorance, where the probabilities are not known, we can base decisions purely on utilities via dominance reasoning, maximin rule or minimax regret.[25] So it seems that the prosecution in *Shonubi III* and the critics of *Shonubi IV* have ignored the one thing that cannot be ignored.

In such cases there is the significant issue of how to determine the relevant utilities. There are, in fact, two quite distinct issues here. The first is that of whose utilities we seek to maximize. After all, the defenses' utilities and the prosecutions' utilities are quite different. Moreover, the utilities of Justice or of society as a whole might be different again. For instance, it does not seem unreasonable to consider the defendant's utilities as being inversely proportional to the size of the sentence, whereas the prosecution's utilities might be thought of as being proportional to the size of the sentence. The second issue is that even once we agree upon whose utilities we wish to maximize, the matter of how we measure them is non-trivial. Do we measure them in terms of years in prison; the natural logarithm of years in prison; financial loss or gain? These two issues are very important and, so far as we can tell, without obvious resolutions. Fortunately we do not need to settle these matters for our purposes here. All

---

[25]Resnik 1987; Jeffrey 1983.

we require is that from any reasonable point of view (for example, from the point of view of Justice) there is significant disutility in handing down a sentence longer than that which is justified.[26] We take it that there is nothing too controversial in this.

Issues relating to sentencing, and therefore to utilities, in the Shonubi case are complex. The sentencing guideline "allows punishment for conduct of which the defendant has not been convicted" if that conduct is deemed relevant to the conduct for which the defendant was convicted (*Shonubi IV*). While the sentencing judge is not required to take such conduct into account s/he is entitled to increase the punishment up to the maximum applicable sentence if appropriate. Furthermore, in the case of Shonubi, the degree of severity of the sentence for the previous seven drug-smuggling episodes, for which Shonubi was neither charged not convicted, could be the same as for the crime for which he was convicted. Since sentencing for drug offenses is determined largely by the quantity of drugs involved, Judge Weinstein deemed it necessary to estimate the aggregate weight of heroin involved across the eight drug-smuggling trips.

Although the appellate court did not specifically appeal to utilities or couch their opinions in decision-theoretic terms, their ruling was clearly in this spirit. They found that:

> [a] guideline system that prescribes punishment for unconvicted conduct at the same level of severity as convicted conduct obviously obliges courts to proceed carefully in determining the standards for establishing whether the relevant conduct has been proven. We have recognized the need for such care with regard to the basic issue of the degree of the burden of proof. Thus, though the Sentencing Commission has favored the preponderance-of-the-evidence standard for resolving all disputed fact issues at sentencing . . . we have ruled that a more rigorous standard should be used in determining disputed aspects of relevant conduct where such conduct, if proven, will significantly enhance a sentence. (*Shonubi IV*)

Hence the appellate court determined that "specific evidence," in the form of drug records, admissions or live testimony, was what was necessary to meet the burden of proof requirements for sentencing Shonubi for the seven prior drug-smuggling trips.

There are two issues at play here. The first is the Second Circuit's concern that evidence that has not brought about a conviction beyond reasonable doubt is used to determine a sentence for relevant conduct at the same level of severity as if the defendant had been convicted with that evidence. The second issue is that since the consequences of using untried evidence are so dire for the defendant (notwithstanding the consequences for legal precedence), the burden of proof relating to the quantities of heroin smuggled by Shonubi in the previous seven trips should be at a higher level than preponderance of evidence. In short, the

---

[26]Of course, the idea of Justice as an agent with a relevant utility function raises all sorts of issues. What is Justice? Can we really think of Justice as an agent? How do we access Justice's utility function?

Second Circuit ruled that since the associated utility plummets with a substantially enhanced sentence (from Shonubi's and the Second Circuit's point of view, in any case), a more rigorous standard of proof is required in the form of "specific evidence." (This is especially so since doubt had been cast on the assumptions behind the statistical analysis.) Fortunately, the Second Circuit considered both utility and probability issues in their final decision.

It is interesting to note that the request by the Second Circuit for a higher standard of evidence was not a request for evidence that implied a higher probability that Shonubi carried over 2,000 grams of heroin. After all, according to the prosecution's statistical analysis, the probability of this was already very high: 0.99. Instead, the request for a higher standard of evidence arose out of concerns about the method employed to derive the figure of 0.99. Thus the Second Circuit was concerned by what we might think of as meta-uncertainty.[27] It is difficult to know how to even quantify such uncertainty let alone provide a convincing case that this uncertainty is low. (For example, how do we answer questions such as: what is the probability that the reference class in question was the appropriate one?) We take the thrust of the Second Circuit's point about utilities to be that in light of the potentially significantly increased sentence involved in this case, we need to be extremely confident that Shonubi carried more than 2,000 grams of heroin over his eight trips. In order to have the required degree of confidence we need both our statistical analysis to yield a high probability for the event in question *and* to have confidence in the method used to obtain that probability. It was the second of these that the Second Circuit was not satisfied with in the Shonubi case.[28]

In closing, we note that if the aggregate estimate of drug quantities for the previous seven occasions had not increased the base offense level of the trip for which Shonubi was actually convicted, there would be no cause for appeal (on the grounds of lack of specific evidence, in any case). It is precisely because of the utilities involved that the sentencing led to appeal.

IV. CONCLUSION

We reject the claim that the Second Circuit Court's findings in *Shonubi IV* amount to a wholesale rejection of statistical evidence in law. Indeed, statistical evidence was used, and accepted by the Second Circuit Court, to estimate the quantity of heroin Shonubi smuggled on the trip for which he was convicted.

[27]This uncertainty has a striking resemblance to what is often called "model uncertainty" in mathematical modelling. When one is using a mathematical model of a complex system to make predictions about that system, no matter how high the probabilities delivered by the model, there remain questions over the reliability of the model itself. More precisely, there is uncertainty surrounding the assumptions on which the model is based and it is notoriously difficult to quantify such uncertainties.

[28]See Mayo (forthcoming) for more on the relationship between meta-uncertainty and decisions with a moral dimension.

Their objection to the statistical evidence presented for the seven prior trips was a cautious and carefully considered reaction to the distinction between the crimes for which Shonubi had been charged and convicted (where there was direct evidence of the quantities involved) and crimes for which he had not been charged (where there was only evidence of the quantities others had smuggled). Moreover, the Second Circuit's rejection of the statistical evidence can be understood as a concern about reference classes, and their suggestion that the preponderance of evidence was too low a standard of evidence in this case can be understood as an appreciation of the importance that utilities play in the problem at hand. Once seen in this light, the Second Circuit's decision is not only, in our view, a fair and just one, but it is also theoretically well motivated.

REFERENCES

Ayer, A. J. 1972. *Probability and Evidence*. New York: Columbia University Press.

Beebee, H. and D. Papineau, D. 1997. Probability as a guide to life. *Journal of Philosophy*, 94, 217–43.

Carnap, R. 1945. The two concepts of probability. *Philosophy and Phenomenological Research*, 5, 513–32.

Cartwright, N. 1983. *How the Laws of Physics Lie*. Oxford: Clarendon Press.

Dowe, P. 1992. A dilemma for objective chance. Paper read to the Australasian Association of Philosophy Conference, University of Queensland, July 1992.

Gastwirth, J. L., B. Freidlin and W. Miao. 2000. The Shonubi case as an example of the legal system's failure to appreciate statistical evidence. Pp. 405–14 in J. L. Gastwirth (ed.), *Statistical Science in the Courtroom*. New York: Springer-Verlag.

Hájek, A. 1997. "Mises redux"—redux: fifteen arguments against finite frequentism. *Erkenntnis*, 45, 209–27.

Hájek, A. Forthcoming. Conditional probability is the guide to life. In H. E. Kyburg and M. Thalos (eds), *Probability as a Guide to Life*. Chicago: Open Court.

Izenman, A. J. 2000a. Assessing the statistical evidence in the Shonubi case. Pp. 415–43 in J. L. Gastwirth (ed.), *Statistical Science in the Courtroom*. New York: Springer-Verlag.

Izenman, A. J. 2000b. Introduction to two views on the Shonubi case. Pp. 393–404 in J. L. Gastwirth (ed.), *Statistical Science in the Courtroom*. New York: Springer-Verlag.

Jeffrey, R. C. 1983. *The Logic of Decision*. Chicago: University of Chicago Press.

Jeffrey, R. C. 1992. "Mises Redux." Pp. 192–202 in Jeffrey, *Probability and the Art of Judgement*. Cambridge: Cambridge University Press.

Kyburg, H. E. 1974. *The Logical Foundations of Statistical Inference*. Dordrecht: Reidel.

Lewis, D. 1986. A subjectivist's guide to objective chance. Pp. 83–113 in Lewis, *Philosophical Papers, Volume II*. Oxford: Oxford University Press.

Mayo, D. G. Forthcoming. Uncertainty and values in the responsible interpretation of risk evidence: beyond clean hands vs. dirty hands. *Risk Analysis*.

Mellor, D. H. 1971. *The Matter of Chance*. New York: Cambridge University Press.

Papineau, D. 1995. Probabilities and the many-minds interpretation of quantum mechanics. *Analysis*, 55, 239–46.

Resnik, M. D. 1987. *Choices: An Introduction to Decision Theory*. Minneapolis: University of Minnesota Press.

Reichenbach, H. 1949. *The Theory of Probability: An Inquiry into the Logical and Mathematical Foundations of the Calculus of Probability*. Berkeley: University of California Press.

Tillers, P. 1997. Introduction: three contributions to three important problems in evidence scholarship. *Cardozo Law Review*, 18, 1875–89.

Tillers, P. 1998. *United States v. Shonubi*: a statistical oddity? http://www.tiac.net/users/tillers/shonubi.html.

Tribe, L. H. 1971. Trial by mathematics: precision and ritual in the legal process. *Harvard Law Review*, 84, 1329–93.

van Fraassen, B. C. 1980. *The Scientific Image*. Oxford: Clarendon.

von Mises, R. 1957. *Probability Truth and Statistics*. 2nd edn. London: Macmillan.

Cases Cited

United States v. Shonubi: *Shonubi V*: 962 F.Supp. 370 (E.D.N.Y. 1997); *Shonubi IV*: 103 F.3d 1085 (2d Cir. 1997); *Shonubi III*: 895 F.Supp. 460 (E.D.N.Y. 1995); *Shonubi II*: 998 F.2d 84 (2d Cir. 1993); *Shonubi I*: 802 F.Supp. 859 (E.D.N.Y. 1992).