

## Legal Decisions and the Reference-Class Problem

Mark Colyvan\* and Helen M. Regan†

There has been a long history of discussion on the usefulness of formal methods in legal settings.<sup>1</sup> Some of the recent debate has focussed on foundational issues in statistics, in particular, how the reference-class problem affects legal decisions based on certain types of statistical evidence.<sup>2</sup> Here we examine aspects of this debate, stressing why the reference-class problem presents serious difficulties for the kinds of statistical inferences under consideration and the relevance of this for the use of statistics in the courtroom. We also consider the relevance of foundational statistical issues in the broader context of

---

\* Professor of Philosophy, Department of Philosophy, A14 Main Quadrangle, University of Sydney, Sydney, NSW, 2006, Australia. *Email:* [mcolyvan@usyd.edu.au](mailto:mcolyvan@usyd.edu.au). I would like to thank members of the TC Beirne School of Law at the University of Queensland, where an early draft of this paper was presented, for their many valuable comments and suggestions. I am also indebted to Jason Grossman and Lena Wahlberg for useful discussions on the topic of this paper.

† Assistant Professor of Biology, Department of Biology, University of California, Riverside, CA 92521, USA. *Email:* [helen.regan@ucr.edu](mailto:helen.regan@ucr.edu).

<sup>1</sup> For example, see L. H. Tribe, ‘Trial by Mathematics: Precision and Ritual in the Legal Process’ (1971) 84 *Harvard Law Review* 1329, and, more recently, R. W. Wright, ‘Causation, Responsibility, Risk, Probability, Naked Statistics, and Proof: Pruning the Bramble Bush by Clarifying the Concepts’ (1998) 73 *Iowa Law Review* 1001, J. L. Gastwirth, *Statistical Science in the Courtroom* (Springer-Verlag, New York, 2000) and P. Tillers, ‘Introduction: Three Contributions to Three Important Problems in Evidence Scholarship’ (1997) 18 *Cardozo Law Review* 1875.

<sup>2</sup> See M. Colyvan, H. M. Regan, and S. Ferson, ‘Is it a Crime to Belong to a Reference Class?’ (2001) 9 *The Journal of Political Philosophy* 168 (reprinted in H. E. Kyburg and M. Thalos (eds), *Probability is the Very Guide of Life* (Open Court, Chicago, 2003)), P. Tillers, ‘If Wishes Were Horses: Discursive Comments on Attempts to Prevent Individuals from Being Unfairly Burdened by their Reference Classes’ (2005) 4 *Law, Probability, and Risk* 33 and R. J. Allen and M. S. Pardo, ‘The Problematic Value of Mathematical Models of Evidence’ (2007) 36 *Journal of Legal Studies* 107.

formal decision theory. In particular, we are interested in cases where the stakes are high — as they are in many legal decisions — and how uncertainty resulting from foundational problems in statistics impacts upon such decisions.

## 1. The Reference-Class Problem

Let us start with an example illustrating the reference-class problem. Consider the population of people in the USA. We wish to know the probability of a given person contracting a particular form of cancer, lung cancer, say. Partition the population into males and females and we get one answer for the probability in question. Partition again into smokers and non-smokers and we get a different answer. Partition again into those exposed to asbestos and those not so exposed, or by geographic region and so on, and we typically get different answers each time. We can even partition into seemingly irrelevant classes, such as people with more than one vowel in their surname and those with only one vowel, and we will get another answer. Despite some of these reference classes seeming to be relevant and some not, there is no principled way to establish the relevance of a reference class. And even if we were able to settle that issue, there is typically more than one relevant reference class (smokers and geographic region, for instance, in the above example). The best we can do, it seems, is to give conditional probabilities as partial answers to our original question. Instead of providing the probability of the individual in question getting lung cancer, we instead provide conditional probabilities for the individual in question contracting lung cancer, *given that the individual in question is male/female, smoker/non-smoker, etc.*<sup>3</sup>

How is this relevant to the use of statistical methods (and formal methods generally) in legal decision making? To see the connection, we revisit the case of *Shonubi*.<sup>4</sup> In this

---

<sup>3</sup> A. Hájek, 'The Reference Class Problem is Your Problem Too' (2007) 156 *Synthese* 563.

<sup>4</sup> *United States v. Shonubi*, 962 F. Supp. 370 (E.D.N.Y. 1997); *United States v. Shonubi*, 103 F.3d 1085 (2d Cr. 1997); *United States v. Shonubi*, 895 F.Supp. 460 (E.D.N.Y. 1995); *United*

case, Nigerian citizen Charles Shonubi was convicted of smuggling heroin into Kennedy Airport in New York. It was determined that he had made seven drug-smuggling trips prior to his arrest. Sentencing for such crimes is, in part, determined by the total quantity of drugs smuggled. The District Court interpreted this to include Shonubi's prior trips, so sentenced him based on the total quantity of drugs in his possession on the trip when he was apprehended, combined with an estimate of the quantity of heroin smuggled on the seven prior trips for which he had not been formally charged. The estimate in question was arrived at via statistical techniques applied to data derived from other heroin smugglers. Ultimately, the US Second Circuit Court of Appeals did not allow the statistical evidence in question and so, in the end, Shonubi was sentenced based only on the quantity of drugs in his possession when apprehended.<sup>5</sup> The debate continues over the significance of the Second Circuit's ruling in this case. Did its ruling amount to a wholesale rejection of statistical evidence in the court room? Did the Second Circuit fail to understand the statistics presented?<sup>6</sup>

In an earlier paper on this topic,<sup>7</sup> we argued that: (i) the reference-class problem raises serious difficulties for the reliability of the estimated probability that Shonubi smuggled specific quantities of heroin on the seven prior trips; and (ii) this uncertainty has ramifications when we consider the problem in its proper context: *as a decision theory problem*. Let us say a little about each of these points, starting with the statistical issues.

---

States v. Shonubi, 998 f.2d 84 (2d Cir. 1993); United States v. Shonubi, 802 F.Supp. 859 (E.D.N.Y. 1992).

<sup>5</sup> Even here the quantity in question was estimated by some rather crude statistical techniques.

<sup>6</sup> See several of the articles in J. L. Gastwirth op. cit., especially A. J. Izenman, 'Introduction to Two Views on the Shonubi Case' in J. L. Gastwirth (ed.) *Statistical Science in the Courtroom* (Springer-Verlag, New York, 2000) 393, for further details of the case and the statistical techniques used.

<sup>7</sup> M. Colyvan, H. M. Regan, and S. Ferson, 'Is it a Crime to Belong to a Reference Class?' (2001) 9 *The Journal of Political Philosophy* 168 (reprinted in H. E. Kyburg and M. Thalos (eds), *Probability is the Very Guide of Life* (Open Court, Chicago, 2003)).

In its estimate of the total quantity of drugs smuggled by Shonubi, the District Court<sup>8</sup> used data on the quantity of heroin smuggled by other Nigerians flying into Kennedy Airport, employing the same mode of transportation (ingested balloons containing a heroin paste mixture), and during the period from Shonubi's first drug smuggling trip until his last. It was assumed that this reference class was appropriate and it was determined there was a 0.99 probability that Shonubi carried at least 2,090.2 grams of heroin on his seven prior drug smuggling trips. Here is where the reference-class problem arises: *other reference classes will give conflicting probability assignments*. Indeed, to use an example from our earlier paper, consider the reference class of toll-booth operators on the George Washington Bridge (of which Charles Shonubi was a member). If one were to base the estimate of the total quantity of drugs Shonubi imported on *this* reference class, presumably there would be a very low probability that he carried at least 2,090.2 grams of heroin on his seven prior drug smuggling trips. Most toll-booth operators, again presumably, do not carry *any* ingested heroin with them when they re-enter the USA from trips abroad. Note that the issue is not about uncertainty related to whether Shonubi is a member of the reference classes in question — he is certainly a member of both classes, and many more besides. Rather, the issue is that, given the different reference classes provide different answers to the probability assignment in question, there is considerable uncertainty about the probability assignment itself.

To put the point slightly differently, the probability of 0.99 assigned to Shonubi carrying at least 2,090.2 grams of heroin on his seven prior drug smuggling trips is a feature of the statistical model used. Other statistical models will give different probability assignments. We thus need either to defend the statistical model used against all others (and we see no prospect for the success of this option) or else we need to live with the resulting meta-uncertainty (uncertainty about the statistical model). This brings us to the second part of our original argument: the importance of decision theory in the face of such meta-uncertainty.

---

<sup>8</sup> United States v. Shonubi, 895 F. Supp. 460, 504–13, 521–23 (E.D.N.Y. 1995), rev'd, 103 F.3d 1085 (2d Cir. 1997).

## 2. Decision Theory and Statistics

We must not forget that statistics, at best, provide justification for belief. Statistical information does not, on its own, motivate action. The question of appropriate action takes us beyond statistics and into the purview of decision theory. Statistics very often provide the probability assignments required for a decision problem but before the correct action can be decided, we also need to consider utilities. Once the probabilities and utilities have been provided, the expected utility is calculated and we are counselled to choose the action associated with the greatest expected utility (if there is such an action). The utilities play a crucial role. To focus exclusively on statistics in a decision problem is to ignore half of what is important. And crucially, for present purposes, the sentencing in *Shonubi* and other legal cases is not a purely statistical problem; it is a decision-theory problem. This important aspect of decision making, legal or otherwise, has been overlooked by some commentators. Our suggestion<sup>9</sup> is that, given the uncertainty over the probability assignment (due to reference-class problems) and the fact that the sentencing in this case (with its novel interpretation of the sentencing guidelines) may involve considerably harsher penalties, the cost of error needs to be considered explicitly. The point is a simple one: when the stakes are low, we can live with extensive uncertainty, but when the stakes are high, we demand higher standards of evidence (and consequently less uncertainty).

This basic piece of decision theory lies behind the different standards of evidence usually involved in conviction (beyond reasonable doubt) and sentencing (preponderance of evidence). The stakes are usually higher in the conviction stage so we require the highest standard of evidence, whereas at sentencing the stakes are typically lower and so lower standards of evidence are acceptable. But the *Shonubi* case is not typical. Here the difference in sentence resulting from the court's interpretation of 'total quantity of drugs' to include trips for which there were no convictions was considerable (about 4.5 years of

---

<sup>9</sup> Previously advanced in M. Colyvan, H. M. Regan, and S. Ferson, 'Is it a Crime to Belong to a Reference Class?' (2001) 9 *The Journal of Political Philosophy* 168 at 176–179.

jail time). This combined with the aforementioned meta-uncertainty, should result in a more precautionary attitude at sentencing. The District Court understated the level of uncertainty by failing to quantify the meta-uncertainty associated with the statistical model used. Given the high disutility of handing down an extra 4.5 year sentence, the Second Circuit appellate bench was right to demand a higher standard of evidence.

In our earlier paper we offered this argument as an — admittedly charitable — interpretation of the Second Circuit’s ruling. The second circuit court did not put the case in these terms. Instead, it demanded ‘specific evidence’ without providing much by way of guidance on what that might consist of. Here we will not dwell on interpretations of the appellate court’s decision any further. In the rest of this article we will defend our argument and comment on recent contributions to the debate by Peter Tillers<sup>10</sup> and Ronald Allen and Michael Pardo.<sup>11</sup>

### **3. Modal Fallacies, Liberal Sympathies and Inferential Arrows**

Peter Tillers raises a number of interesting points in relation to our arguments. First, he suggests that we have fallen for a modal fallacy: believing that  $P$  follows from the modal proposition ‘I wish that  $P$ ’.<sup>12</sup> The suggestion is that we would like it to be the case that all

---

<sup>10</sup> P. Tillers, ‘If Wishes Were Horses: Discursive Comments on Attempts to Prevent Individuals from Being Unfairly Burdened by their Reference Classes’ (2005) 4 *Law, Probability, and Risk* 33.

<sup>11</sup> R. J. Allen and M. S. Pardo, ‘The Problematic Value of Mathematical Models of Evidence’ (2007) 36 *Journal of Legal Studies* 107.

<sup>12</sup> We put aside Tillers’ (‘If Wishes were Horses’, 37) claim that we argued that the central claim of P. Tillers, ‘Introduction: Three Contributions to Three Important Problems in Evidence Scholarship’ (1997) 18 *Cardozo Law Review* 1875 was ‘nonsense’. We claimed no such thing. We argued against his position; we could hardly do this if his position were nonsense (i.e. without meaning). Falsity and nonsense are very different. Nor will we pursue his claim (‘If Wishes Were

good evidence is specific and from this we conclude that all good evidence is specific. But this is not right. What we would like to be the case never enters into it. We *argue* that the reference-class problem casts doubt on the statistical evidence presented. As it turns out, we wish that it didn't, but as Tillers rightly points out, we don't always get what we wish for. Tillers goes on to suggest that liberal sympathies might be driving all this.

It is possible that [...] liberal and humane sentiment animates or colours the thinking of Judge Newman—and also that of Colyvan, Regan and Ferson—about the importance of 'specific evidence' and the dangers of 'non-specific' evidence. If so, I applaud them all—Newman, Colyvan, Regan and Ferson—for their humane spirit and for their liberal sentiments. Nonetheless, I cannot applaud the suggestion of these caring and well-meaning people that inferences based on non-specific evidence are always inferentially unsound.<sup>13</sup>

Nowhere in our discussion of reference classes do such sympathies play a role. Indeed, we deliberately turn things around and give the example of using the reference class of toll-booth operators on the George Washington Bridge as a means of establishing Shonubi's innocence, and we point out that that would be a decidedly bad defence. The bottom line is that the argument in question is statistical, not social or political.

Although Tillers does not discuss our comments on decision theory, his charge against us might have some force there. For it is in the decision-theory setting that utilities enter into the picture, and it is here that more liberal sympathies might well play a role.<sup>14</sup> After all, it is liberal sympathies that drive the different standards of proof in court and these same liberal sympathies drive the presumption of innocence. The idea is that Type I errors — false positives — are more serious (at least in the legal setting) than Type II errors —

---

Horses', 38) that we took ourselves to have solved the reference-class problem. We state quite explicitly in our paper that we do not claim to have done this.

<sup>13</sup> Tillers, 'If Wishes were Horses', 45.

<sup>14</sup> See M. Colyvan, D. Cox, and K. Steele, 'Modelling the Moral Dimension of Decisions' (to appear) for ways of explicitly introducing ethical consideration into the decision-theoretic framework.

false negatives. That is, convicting someone who is innocent — a false positive — is worse than failing to convict someone who is guilty — a false negative. We suggested that because the stakes are so high in sentencing decisions like the one in *Shonubi*, and in light of the metauncertainty resulting from the reference-class problem, higher standards of proof should be required than the orthodox preponderance standard. To suggest otherwise would seem to fly in the face of liberal legal tradition.

The more substantial matter raised by Tillers is whether one can avoid what he calls ‘inferential arrows’:

[I]t is a mistake to suppose that any individual can escape the inferential arrows that the natural and human evidential environment inevitably throws her way, and it is then an error for an individual to think he, she or any individual can escape being subjected to inferences that depend on observations and judgements about the behaviour and attributes of other human beings. Not in a pig’s eye! No one can altogether avoid signs, signposts and evidentiary hints that the operations of the world and other people have created. It is the failure to appreciate this basic point that makes the position of Judge Newman, Colyvan, Regan and Ferson unsatisfactory and, ultimately, untrue.<sup>15</sup>

Tillers’ suggestion here is that it is always possible to draw conclusions about an individual’s behaviour based on evidence gathered by others. We agree. The problem is that this evidence will depend on the reference class used. There are simply too many evidential arrows and they don’t always point in the same direction! Tillers takes us to be denying that it is ever legitimate to engage in group to individual inferences:

It is true that many inferences about individual behaviour based on statistics about group behaviour or about other people are just rotten inferences. But it hardly follows that all inferences about individuals based on statistics about groups are inferentially worthless.<sup>16</sup>

---

<sup>15</sup> Tillers, ‘If Wishes were Horses’, 49.

<sup>16</sup> Ibid 44.

But we agree with him that not all inferences about individuals, based on statistics about groups, are worthless. Consider the inference all *F*s are *G*s, *a* is an *F*, therefore *a* is *G*. This is most certainly *not* a ‘rotten inference’; to the contrary, it is deductively valid.<sup>17</sup> But Tillers, quite rightly, is not interested in such limiting cases of statistical arguments. His claim is that there are good arguments based on (non-trivial) statistics. Again we agree. Indeed, we went to considerable lengths in our earlier paper to say why we did not think that the rejection of the statistical evidence in question in *Shonubi* amounts to the wholesale rejection of statistical methods in the courtroom.<sup>18</sup> Such wholesale rejection would only follow if the statistics in this case were as good as it gets. But as we have argued, the issue of meta-uncertainty was ignored, so things most certainly were not as good as it gets in *Shonubi*. Tillers takes us to be challenging his claim that ‘it is *not* always inferentially illegitimate to base inferences about the behaviour of an individual on the behaviour of other people’.<sup>19</sup> But we took our target to be a quite different thesis. We took our target to be the thesis that in the *Shonubi* case, the statistical analysis *as it stood* supported the inference about the behaviour of Shonubi from the behaviour of the others in the reference class in question. Now of course our argument against this thesis proceeded via a general discussion of the reference-class problem, and this is (according to us) a wide-spread problem. But at no stage did we suggest that it is always a mistake to

---

<sup>17</sup> Tillers uses the term ‘valid’ in a looser sense, for example, when he asserts that ‘valid inferences are unfair’ (‘If Wishes Were Horses’, 46). Here, we take it, that rather than validity, he is really talking about good inductive inferences. After all, none of the inferences under discussion is even a candidate for (deductive) validity; all non-trivial statistical arguments are deductively invalid.

<sup>18</sup> For example, various forms of sensitivity analysis would help bolster the statistical inferences in question. For example,, if the same or similar results were obtained across a broad class of statistical models, the meta-uncertainty is reduced and we can have more confidence in the inferences drawn from a particular statistical model. Admittedly, such cases may be rare, but that’s not the issue. Such methods of tackling meta-uncertainty are available and they were not implemented in the statistical analyses in the *Shonubi* case.

<sup>19</sup> Tillers, ‘If Wishes were Horses’, 39.

draw inferences about an individual based on statistical arguments from a population.<sup>20</sup> Our point is that in drawing such inferences one must accept, and deal with, the resulting uncertainty. In *Shonubi*, this uncertainty was obscured, indeed understated, by sweeping the reference-class problem aside.<sup>21</sup>

#### **4. Formal Methods and Inference to the Best Explanation**

Allen and Pardo also take up the issue of the reference-class problem in legal decision making.<sup>22</sup> We find ourselves in broad agreement with most of what they have to say. We do not, however, share their pessimism about the use of formal methods.

Allen and Pardo suggest that the lesson of the *Shonubi* case and subsequent debate is that ‘inferential problems at trial appear to defy formal treatment’. They come to this conclusion for several reasons. The first is the meta-uncertainty (uncertainty about the appropriate choice of reference classes, for example) that we drew attention to in our earlier paper and which we have discussed again here. A second reason for Allen and Pardo’s pessimism over the prospects for formal treatments of legal inferences is the

---

<sup>20</sup> Tillers again puts the point in terms of validity and invalidity (‘If Wishes Were Horses’, 40), but in the technical usage of this term, at least, all parties agree that the inferences in question are invalid; the issue is whether the inferences in question are good inductive inferences.

<sup>21</sup> Tillers proceeds to argue that we ‘ignored the elephant in the room’ by failing to examine the Second Circuit’s reference to ‘specific evidence’ and its failure to distinguish between evidentiary and socio-political considerations. But this is puzzling, since both issues were directly addressed in our original commentary: see ‘M. Colyvan, H. M. Regan and S. Ferson ‘Is it a Crime to Belong to a Reference Class?’’, at 175–176 and 176–179 respectively.

<sup>22</sup> R. J. Allen and M. S. Pardo, ‘The Problematic Value of Mathematical Models of Evidence’ (2007) 36 *Journal of Legal Studies* 107.

conspicuous lack of clarity surrounding key terms such as ‘more reliable’ and the like.<sup>23</sup> And it is this that seems to be their principal reason for pessimism:

This can be seen by flipping the *Shonubi* question and asking of the District Judge or his academic assistants to explain why the evidence actually provided at trial and sentencing established by a preponderance of the evidence that Shonubi imported a certain amount of drugs. No formal answer will be forthcoming to that question, and the answer ultimately will be in the form of the exercise of judgment. Precisely so, as always will be the case at trial, we suspect. Thus, the dispute between the trial and appellate courts was not a profound disagreement about probability theory or statistics but instead about the exercise of judgment over the facts of this case.<sup>24</sup>

From the discussion leading up to this passage it is clear that it is vagueness in some of the key legal terms that is the issue. Indeed, vagueness presents serious difficulties for classical logic and hence classical probability theory.<sup>25</sup> But classical probability theory is not the only game in town. There are various non-classical probability theories and these are much better suited to representing uncertainty due to vagueness. For example, there

---

<sup>23</sup> Allen and Pardo (ibid) suggest that the problem is ambiguity in the terms. But what they are really complaining about, we take it, is vagueness. Ambiguity is where a term has more than one (well defined) meaning. Contrast this with vagueness, which is where the term in question allows borderline cases (see H. M. Regan, M. Colyvan, and M. A. Burgman. ‘A Taxonomy and Treatment of Uncertainty for Ecology and Conservation Biology’ (2002) 12 *Ecological Applications*, 618). An inference can be more or less ‘more reliable’, which makes it vague; it is not that ‘more reliable’ means more than one thing. We all agree what ‘more reliable’ means (at least in this context), it’s just that we don’t know where to draw the line or what to do if we do not draw a sharp line between more reliable and less reliable. See also L. Laudan, *Truth, Error, and Criminal Law: An Essay in Legal Epistemology* (Cambridge University Press, Cambridge, 2006) for discussion of the unclarity of other, key legal terms.

<sup>24</sup> Allen, and Pardo ‘The Problematic Value of Mathematical Models of Evidence’, 135.

<sup>25</sup> Classical probability theory presupposes classical logic: Kolmogorov probability theory assigns maximal probability to the tautologies of *classical logic* (see M. Colyvan, ‘The Philosophical Significance of Cox’s Theorem’ (2004) 37 *International Journal of Approximate Reasoning* 71).

are various *vague probability theories*, which avoid the rather stringent requirement that belief be modelled by a single real-valued probability function.<sup>26</sup> We are not claiming that such treatments are straightforward or that they solve all the problems at hand. On the contrary, a great deal of work in philosophical logic continues on the many technical, practical and philosophical problems associated with vagueness. Our point is simply that important work is being done and progress is being made. Moreover, this work is formal. It involves developing formal logical and probability theories that handle vagueness better than their classical counterparts. So while we agree with Allen and Pardo that vagueness presents serious obstacles to the use of formal methods in legal decision-making, this should not be thought to be a reason for giving up on formal methods *per se*, though it may well be a reason to give up on classical theories.

In place of formal methods Allen and Pardo propose<sup>27</sup> invoking inference to the best explanation as a way forward. They rightly point out that explanation is important in the kinds of inferences salient in legal settings, and yet explanation is by and large ignored by standard statistical analyses. An example will help. When it was found that Shonubi had multiple passports, and had made numerous overseas trips that would have been too expensive to fund on a toll-booth operator's wages, it was inferred that several of the trips in question involved illegal transportation of drugs. Indeed, something like this form of reasoning was invoked at the District Court in concluding that Shonubi had made seven prior drug-smuggling trips. Notice that there is no appeal to the frequency of drug-runners in the population of multiple passport-holders who engage in overseas travel beyond their means. The hypothesis that the seven trips in question were drug-smuggling trips, best explains the evidence in question, and that is enough. Or so the story goes.

---

<sup>26</sup> Non-classical theories in this spirit have been advanced by H. Field, 'Indeterminacy, Degree of Belief and Excluded Middle' (2000) 34 *Noûs* 34 1, G. Shafer, *A Mathematical Theory of Evidence* (Princeton University Press, Princeton, NJ, 1976) and P. Walley, *Statistical Reasoning with Imprecise Probabilities* (Chapman and Hall, London, 1991).

<sup>27</sup> Following our earlier, somewhat tentative, suggestion in M. Colyvan, H. M. Regan, and S. Ferson, 'Is it a Crime to Belong to a Reference Class?' (2001) 9 *The Journal of Political Philosophy* 168 at 175–176.

While it is clear that inference to the best explanation is widely used in both scientific and every-day inferences,<sup>28</sup> it is not clear that it eschews formal methods. After all, just because the inference that Shonubi made seven prior drug-smuggling trips was not based on frequency data, this does not mean that it did not involve probabilities. In our earlier paper we suggested that inferences such as these might involve subjective probabilities.<sup>29</sup> These might be based on other subjective probabilities such as the probability that drug smugglers make multiple trips, the probability that a known drug smuggler would have other legal but invisible income, and so on. Allowing subjective probabilities in place of statistical frequencies is not, in our book, giving up on formal methods.

We also note that inference to the best explanation does not seem applicable unless there is already quite a bit of evidence available. After all, evidence that Shonubi was a drug smuggler and that drug smugglers are unlikely to have other, legal means of living beyond their visible means plays an important role in the courts' reasoning in *Shonubi*. To see this, consider how poor the inference is that a randomly-chosen person with a low-paying job, multiple passports and regular overseas trips is a drug runner. To put the point slightly differently, when the evidence is thin it may be very difficult to single out a privileged 'best' explanation, or worse still, the best explanation might be clearly unacceptable (e.g. when only one explanation is on offer). The bottom line is that sometimes we need to refrain from endorsing the best explanation.<sup>30</sup> Moreover, we suspect that such situations arise in many legal settings. This is not to suggest that inference to the best explanation is not useful in such settings, just that one needs to be careful with its application and aware of the difficulties associated with this controversial form of inference. But these cautions aside, Allen's and Pardo's suggestion is a good one and deserves more discussion than we have been able to provide here.

---

<sup>28</sup> Although it is by no means uncontroversial. For a very good critical discussion, see B. C. van Fraassen, *The Scientific Image* (Clarendon, Oxford, 1980).

<sup>29</sup> M. Colyvan, H. M. Regan, and S. Ferson, 'Is it a Crime to Belong to a Reference Class?' (2001) 9 *The Journal of Political Philosophy* 168 at 176.

<sup>30</sup> We put aside the vexed issue of how to settle disputes over what the best explanation is.

## 5. Conclusion

In conclusion, we would like to say a little about the role of reference-class evidence — evidence gathered from others (or what Tillers calls the inferential arrows) — because this is really the heart of the reference-class problem. Again, we rely on decision theory to illuminate the appropriate use of such evidence in the course of establishing guilt. We have argued that the reference-class evidence used in Shonubi's sentencing was inappropriate in light of the disutility (to society at large) of inflicting undeserved, additional prison time on Shonubi. This would not be an issue if Shonubi's guilt in smuggling a total of 2,090.2 grams of heroin were certain. The disutility becomes important when we consider the uncertainty associated with the choice of reference class on which the sentencing decision relies. Our argument is that when the probabilities themselves are uncertain (because of uncertainty over the appropriate reference class), we should proceed decision theoretically, and this involves explicitly invoking utilities. Serious consequences demand greater degrees of epistemological warrant. On the other hand, when the stakes are low we can tolerate much higher levels of uncertainty because the consequences of being wrong aren't so grave. Either way, the utilities help determine the appropriate standard of evidence.

Uncertainty will always be prevalent in legal decisions. Mathematical models of evidence provide some relief from uncertainty but can suffer from reference-class problems. This introduces a source of meta-uncertainty, which may well be impossible to eliminate. Decision theory, however, is in the business of recommending the best course of action in the face of uncertainty and thus can provide a valuable framework for approaching legal decisions and their associated uncertainties.