Comments on Jim Franklin's "The Representation of Context: Ideas from Artificial Intelligence"

(Or, More Remarks on the Contextuality of Probability)

Branden Fitelson

April 28, 2003

To be honest, I have almost nothing critical to say about Jim's presentation (and this is quite unusual for a cranky analytic philosopher like me!). What Jim has said is all very sensible, and his examples are very well chosen, *etc.* So, instead of making critical remarks, I will try to expand a little on one of the themes Jim briefly touched upon in his talk: the contextuality of probability.

Jim mentioned that personalistic Bayesian probabilists run squarely into the problem of contextuality every time they apply Bayes' Theorem. When a Bayesian asks "What should my degree of belief in H be (*i.e.*, what *probability* should I assign to H), given my total evidence E?", they will have to be able to say what the probability of H should have been *prior to* having learned E. These "prior probabilities" are only scrutable in the context of the entire structure of the agent's prior beliefs (Bayesian belief structures are assumed to have a fair amount of *a priori* structure, *e.g.*, Boolean algebra). This contextuality of subjective probabilities is well known, and has been a source of great controversy and pain for practicing Bayesians.

What I would like to emphasize in my brief remarks is that *many* kinds of probability face equally deep problems of contextuality: this is not just a problem for Bayesians (*i.e.*, *subjectivists* about probability). I will illustrate this point with two examples. First, an example involving an infamous contemporary legal case. After the O.J. trial, Alan Dershowitz remarked on T.V. (this was quoted widely in the papers — see [2]) that "fewer than 1 in 1,000 women who are abused by their mates go on to be killed by them". Here, Dershowitz is not talking about a subjective probability — he is talking about an *objective* probability (a *frequentist* or *statistical* probability, I presume). And, given the context, we may assume that he intended this remark to be relevant to the estimation of the probability of a *particular* abuser's guilt of murder (O.J.).

To fix our ideas, and to make things more precise, let A be the proposition that Nicole Brown Simpson is abused by her spouse (O.J.), let K be the proposition that Nicole Brown Simpson is killed by her spouse (or ex-spouse, O.J.), and let $\Pr(\cdot\,|\,\cdot)$ be whatever objective (say, statistical) conditional probability function Dershowitz has in mind here (more on this below). Now, Dershowitz is claiming (roughly) that:

(1) $$\Pr(K\,|\,A) < \frac{1}{1000}$$

Shortly after Dershowitz made this remark, the great statistician I.J. Good wrote a brief response in *Nature* [5]. Good pointed out that, while Dershowitz's claim may be true, it is not salient to the case at hand, since it ignores crucial additional contextual information that is available to us. Good claims that what's *really* important here is not the probability that Nicole Brown Simpson was killed by O.J., given that she was abused by O.J., but the probability that she was killed by O.J., given that she was abused by O.J. *and that she was killed.* After all, we do know that Nicole was killed, and (plausibly) this information should be taken into account. Let K' be the proposition that Nicole Brown Simpson was killed (by *someone*). Using Dershowitz's (1) as a starting point, Good does some back-of-the-envelope calculations (largely on the basis of *speculation* about the other relevant frequencies), and comes up with the following "guesstimate":

(2) $$\Pr(K\,|\,A\,\&\,K') \approx \frac{1}{2} \gg \frac{1}{1000}$$

This would make it far more probable that O.J. is the killer than Dershowitz's claim would have us believe. Independently, but using *actual data* about murders committed in 1992, Merz & Caulkins [6] estimate that:

(3) $$\Pr(K\,|\,A\,\&\,K') \approx \frac{4}{5}$$

This would seem to provide us with an *even greater* probability of O.J.'s guilt. Dershowitz [2] replies to analyses like those of Good and Merz & Caulkins by pointing out that (my brackets) "... whenever a woman is murdered, it is highly likely that her husband or her boyfriend is the murderer without regard to whether battery preceded the murder. The key question is how salient a characteristic is the battery as compared with the relationship itself. Without that information, the 80 percent figure [as in Merz & Caulkins' estimation] is meaningless. I would expect that a couple of statisticians would have spotted this fallacy." Using our notation, Dershowitz's rejoinder seems to trade on something like the following claim:

(4) $$\Pr(K \mid K') \approx \Pr(K \mid A \,\&\, K') \quad [\textit{i.e.}, K', \text{ not } A, \text{ is doing the real work here}]$$

Not to be outdone, Merz & Caulkins estimate that (again, based on actual murder statistics for 1992):

(5) $$\Pr(K \mid K') \approx 0.29 \ll \Pr(K \mid A \,\&\, K') \approx 0.8$$

We could continue this dialectic *ad nauseam*. I'll let you add whatever further epicycles you like to the opposing sides of this dispute. Probabilists (and philosophers of probability) are all to familiar with *this* morass! This is often called "the reference class problem".[1] That's just a fancy name for *the contextuality of objective probabilities* (in this case, *frequencies* or *statistical* probabilities).

Here's a more striking example illustrating the contextuality of statistical probabilities. Question: "What's the probability that John Doe has cancer?" Answer: there are *many* ways to answer this question, depending on which reference class John Doe is included in. It turns out that John is 65 years old. If we just look at the frequency of cancer among 65 year-old males, then (presumably) we'll get a relatively large (at least, non-trivial) number. But, John is also an avid runner and a vegetarian. Presumably, this bumps the number down. However, he's a long-time smoker, which (say) bumps the number back up again, *etc.* This sequence can be continued in various ways. So, which reference class is "the right one"? Which context will give us "the true probability that John Doe has cancer" (analogy with special relativity: which frame of reference gives "the true velocities"?)? Perhaps we should include *all* of the properties that John has, in order to determine "the true probability". This will not do, since it will *uniquely* pick-out John Doe, which will force the probability in question (understood as a "frequency" of cancer in the singleton "reference class" {John Doe}) to be either one or zero (depending on whether or not John Doe in fact has cancer).

I will not try to offer a solution to this problem here (my Nobel Prize awaits me when I do!). I aim here only to expand upon what Jim has said about the contextuality of subjective probability. The main lesson is that moving away from subjective probability to a more objective variety (like frequencies or statistical probabilities) does not eliminate the contextuality problem — it seems that probability claims of *any* kind are only meaningful and precise *in relation to specific probability models.* I think it is important to emphasize this more general point, since some people might have the impression that the contextuality of (personalistic) Bayesian probability stems only from its subjectivity. This would be a false impression.[2]

References

[1] M. Colyvan, H. Regan, and S. Ferson. Is it a crime to belong to a reference class? In H.E. Kyburg and M. Thalos, editors, *Probability is the Very Guide of Life*, pages 331–347. Open Court, 2003.

[2] A. Dershowitz. The numbers game: Letter to the editor. *The New York Times*, 30 May 1999.

[3] R. Festa. *Optimum Inductive Methods.* Reidel, 1993.

[4] D. Gillies. *Philosophical Theories of Probability.* Routledge, 2000.

[1] Reichenbach [7] was one of the first people to write carefully about this problem, and since then there have been many attempts to resolve it. For an excellent contemporary article on the reference class problem in legal contexts, see [1].

[2] Indeed, this problem seems to plague *all* interpretations and accounts of probability. Here are some further examples, with references. The Bertrand paradoxes [8, chapter 12] show that *Classical* probabilities face a contextuality problem (contextual factors seem crucial for the determination of "the correct equipossible partition"). *Propensity* theories of probability [4] are unabashedly contextual (the "experimental set-ups" that fix the propensities are just gussied-up contexts). Even *Logical* (or *Inductive*) probabilities seem to require appeals to context (all attempts to formulate a "context-free" conception of logical or inductive probability have failed pretty miserably — see [3] for a compelling and explicitly contextual approach).

[5] I.J. Good. When batterer turns murderer. *Nature*, page 541, 15 June 1995.

[6] J. Merz and J. Caulkins. Propensity to abuse—propensity to murder? *Chance Magazine*, page 14, 1995.

[7] H. Reichenbach. *The Theory of Probability. An Inquiry into the Logical and Mathematical Foundations of the Calculus of Probability.* University of California Press, second edition, 1971.

[8] B. van Fraassen. *Laws and Symmetry.* Oxford University Press, New York, 1989.