Comments on "A foundational justification for a weighted likelihood approach to inference", by Russell J. Bowater

Glenn Shafer
Rutgers Business School
180 University Avenue
Newark, New Jersey 07104
gshafer@rutgers.edu
www.glennshafer.com

December 3, 2003

Russell Bowater uses non-additive probability to interpret weighted likelihood. Because I have written a great deal about non-additive probability, the editor has asked me to comment.

1 Non-Additive Probability

Non-additive probability goes back to the very beginning of probability theory—the work of Jacob Bernoulli. Bernoulli's calculus for combining arguments allowed both sides of a question to attain only small or zero probability, and he also thought the probabilities for two sides might sometimes add to more than one (Shafer 1978).

Twentieth-century non-additive probability has roots in both mathematics and statistics. On the mathematical side, it is natural to generalize measure-theoretic probability by interpreting upper and lower bounds on the measure of a non-measurable set as the set's non-additive "upper and lower probabilities". On the statistical side, it is natural to try to use the greater flexibility of upper and lower probabilities in an effort to find better solutions to problems of inference. A. P. Dempster (1968) and Peter Walley (1991), perhaps the most influential innovators in this domain, both proposed generalizations of Bayesian inference.

In my work on the "Dempster-Shafer theory" in the 1970s and 1980s (Shafer 1976), I called the lower probability ($\underline{P}$ or P^* or Bel) a degree of support or belief. It measures the strength of evidence for an event but does not necessarily have a betting interpretation. The upper probability ($\overline{P}$ or P_* or Pl) I called "plausibility". An event or proposition is plausible to the extent its denial is not supported. The Dempster-Shafer calculus allows one to score and combine arguments in Bernoulli's sense, and it has proven useful in systems where there are many different sources of evidence, all of which merit attention but none of which can be calibrated in terms of frequency or betting.

It is also natural, however, to define upper and lower probabilities in terms of betting, generalizing de Finetti's treatment of additive probability. To my knowledge, the first person to take this path was Robert Fortet, in 1951. Others who followed in subsequent decades include C. A. B. Smith (1965), Peter Williams (1976), and especially Peter Walley (1991). At least since Williams, most of this work has started with the idea of upper and lower previsions (or expectations) for variables rather than with the narrower idea of upper and lower probabilities for events. Walley's work has inspired a small ongoing community that studies "imprecise probabilities" (see Bernard, Seidenfeld, and Zaffalon 2003).

One way to frame the betting definition of upper and lower previsions is to imagine a market (e.g., a stock market) where certain variables are available (perhaps in negative as well as positive quantities) at announced prices. By putting together portfolios of these priced variables, we can obtain upper and lower prices for an arbitrary variable x:

- The *upper prevision* $\overline{E}(x)$ is the lowest price at which the market will sell us x. More precisely, it is the least (or infimum) total cost for a portfolio of priced variables that will pay back at least x no matter what happens.

- Similarly, the *lower prevision* $\underline{E}(x)$ is highest price at which the market will buy x from us. Because buying x for c is the same as selling $-x$ for $-c$, this means $\underline{E}(x) = -\underline{E}(-x)$.

The *upper probability* $\overline{P}(A)$ (resp. *lower probability* $\underline{P}(A)$) for an event A is the upper prevision (resp. *lower prevision*) for A's indicator function.

Vladimir Vovk and I have added an explicit perfect-information sequential game to this picture (Shafer and Vovk 2001). Defining upper and lower previsions using trading strategies rather than mere static portfolios, we then have a framework for pricing options. We also obtain generalizations of the classical limit theorems of probability.

An alternative way of thinking about upper and lower probabilities is to regard them as bounds on probability measures. Many authors have explored this idea, but its complexity seems to make it less natural and useful than the direct betting picture.

1.1 Ways of using the betting picture

The simplest way of relating the betting picture to the real world, perhaps, is to take it as an idealized description (abstracting from trading costs) of financial markets. As I have already suggested, this leads to a useful theory of option pricing.

Another way using the picture leads to a notion of subjective non-additive probability—Bowater's "type II probability". I put myself in the role of the market in the picture, so that $\underline{E}(x)$ is the most I am willing to pay for x, and $\underline{P}(A)$ the most I am willing to pay for the indicator function for A. Bowater is following Walley and many other authors when he uses the picture in this way.

Other ways of using the betting picture arise from what Vovk and I call "Cournot's principle": the market will not allow a trader to become exceedingly rich without risking bankruptcy. We can use this principle in several different ways to relate the betting picture to the real world:

- In the context of an actual securities market, Cournot's principle is an efficient-market hypothesis. It leads to a theory of finance that explains certain constraints on security prices (Vovk and Shafer 2002).

- Suppose that instead of imagining myself in the role of the market (who offers bets), I imagine myself in the role of a trader in the market (who contemplates taking bets). Someone else sets the prices (this might be an individual, an algorithm, or a statistical model), but I adopt them as my subjective previsions by adopting the belief that reality will not allow me to get rich trading at these prices. This leads to a concept of subjective probability somewhat different from Bowater's (Shafer, Gillett, and Scherl 2003).

- Instead of merely asserting a personal belief that I cannot get rich at prices given by some model or theory, I may assert this as a fact about the world. This leads to a concept of objective probability, which Vovk and I argue is appropriate when we are working with a well-tested theory such as quantum mechanics (Shafer and Vovk 2001, §8.4).

Many readers may be familiar with the version of Cournot's principle that applies to classical or measure-theoretic probability: an event of small or zero probability, singled out in advance, will not happen. Many classical probabilists, including Émile Borel, Jacques Hadamard, Paul Lévy, and Andrei Kolmogorov, used this principle to relate probability theory to the real world (Shafer and Vovk 2003). (I do not agree, by the way, that these classical authors were "vague" compared with Bowater and the authors he cites, Ramsey, de Finetti, and Pratt.)

1.2 Bowater's theory of nonadditive probability

As I have already indicated, Bowater's type II probability is what Walley and others in the "imprecise probabilities" community call lower probability. (Had I seen his article in advance, I would have suggested that Bowater signal this by writing $P_*(A)$ rather than $P^*(A)$.) So we should ask what light the work of this community casts on what Bowater does.

Bowater defines conditional type II probability as "the maximum amount the individual would be willing to pay ... given the individual knows that the event B has occurred." Presumably he has in mind the same kind of thought experiment Walley considers: one thinks about how one would change one's beliefs if one learned B were true. If this interpretation is correct, then Bowater's Assumption 2 is a consequence of Walley's updating principle, which has been much studied and debated. My own view, explained in the article with Gillett and Scherl already cited, is that the updating principle is really justified only when we use Cournot's principle.

On the other hand, Bowater's third assumption (that ordinary probability can be taken as type II probability when it exists) is unnecessary in Walley's framework. Walley has a principle of "no sure loss", and this is enough to rule out different probabilities for the same events.

What is distinctive—and probably unjustified—in Bowater's story is his emphasis on atoms. From the viewpoint of Walley's theory, lower probabilities for individual values of a parameter θ are not necessarily very interesting. Mathematically, we can give each individual value lower probability zero while putting substantial lower probability on larger subsets of the parameter space, and sometimes it may be reasonable to do so. Even if the parameter space is discrete, we may be unwilling to bet on any individual value even at very favorable odds.

2 Direct Interpretation of Likelihood

Bowater mentions several authors who have promoted the direct interpretation of likelihood, and we should surely add R. A. Fisher, whose advocacy was original and persuasive to many. The authors on likelihood that Bowater cites added little to Fisher's viewpoint; perhaps this is why Bowater calls their work applied.

So far as I know, Fisher did not talk about weighted likelihood. But he did mention the idea of aggregating evidence from independent sources by multiplying likelihoods (1973, p. 134), and many authors, including A. W. F. Edwards (1972, p. 35), have pointed out that this can be interpreted as a fairly general rationale for weighting a statistical likelihood $l(\tilde{x}, \theta)$ with prior beliefs derived from other evidence. We simply think of the prior evidence as an independent observation, assess its probability under the different values of the parameter θ, and then multiply this "prior likelihood" by $l(\tilde{x}, \theta)$ to obtain a joint likelihood.

What Bowater contributes is a new way of thinking about what we are doing when we interpret the likelihood directly. Is this new way of thinking easier or more persuasive?

It may help us think about Bowater's contribution to divide Fisher's argument into two steps.

1. First, Fisher wanted to persuade us that parameter values should be considered likely in proportion to the probability they give to the observation.

2. Second, he wanted us to be content with a relative concept of likelihood. If we know that a range $[\theta_1, \theta_2]$ includes everything that is at least 1/20th as likely as the most likely value, maybe we know enough about θ. Let's

work on something else. We don't need a "probability" for θ being in the interval $[\theta_1, \theta_2]$.

Both steps are uncomfortable for a person who wants probabilities for θ.

The strength of Bowater's contribution, I think, is that he gives us something more like probability at step 1. He gives us something we can interpret in terms of betting. The relation $P^*(\theta_1|\tilde{x}) = 3P^*(\theta_2|\tilde{x})$ means that we would pay three times as much to get back a pound sterling if θ_1 is true than to get back a pound sterling if θ_2 is true.

The weakness of Bowater's contribution is that he goes along with Fisher on step 2, even though he has adopted what Fisher would have called a stronger "logical type" of inference. Instead of giving us a lower probability for θ being in $[\theta_1, \theta_2]$, he tells us only that $[\theta_1, \theta_2]$ includes all the individual values of θ that have lower probability at least 1/20th as great as the maximum lower probability for an individual value of θ. Is this all we want to know? Bowater's principles suggest that the interval $[\theta_1, \theta_2]$ does have a lower probability, and it is mathematically possible that this lower probability is very small even though the interval includes all the values of θ with high individual lower probability. In fact, we might even have a lower probability close to one that θ is not in $[\theta_1, \theta_2]$.

I suspect that Bowater will not be joined by a large company at the point on the trail where he has paused to rest. Those who prefer the objectivity and simplicity of likelihood will stay back with Fisher. Those who are strongly enough attracted by Bowater's type II probability to trace his steps will likely forge on to explore new non-additive territory.

References

[1] Raymond Bayer, editor. *Congrès International de Philsophie des Sciences, Paris, 1949; IV: Calcul des Probabilités.* Number 1146 in *Actualités Scientifiques et Industrielles.* Hermann, Paris, 1951.

[2] Jean-Marc Bernard, Teddy Seidenfeld, and Marco Zaffalon. *ISIPTA '03. Proceedings of the Third International Symposium on Imprecies Probabilities and their Applications.* Carleton Scientific, Waterloo, Canada, 2003.

[3] A. P. Dempster. A generalization of Bayesian inferenece. *Journal of the Royal Statistical Society, Series B*, 30:205247, 1968.

[4] A. W. F. Edwards. *Likelihood.* Johns Hopkins, Baltimore, 1972. Second edition 1992.

[5] Ronald A. Fisher. *Statistical Methods and Scientific Inference.* MacMillan, New York, third edition, 1973. First edition 1956, second 1959.

[6] Robert M. Fortet. Faut-il élargir les axiomes du calcul des probabilités? In Bayer [1], pages 35–47.

[7] Glenn Shafer. *A Mathematical Theory of Evidence.* Princeton University Press, Princeton, New Jersey, 1976.

[8] Glenn Shafer. Non-additive probabilities in the work of Bernoulli and Lambert. *Archive for History of Exact Sciences*, 19:309–370, 1978.

[9] Glenn Shafer, Peter R. Gillett, and Richard B. Scherl. A new understanding of subjective probability and its generalization to lower and upper prevision. *International Journal of Approximate Reasoning*, 33:1–49, 2003.

[10] Glenn Shafer and Vladimir Vovk. *Probability and Finance: It's Only a Game.* Wiley, New York, 2001.

[11] Glenn Shafer and Vladimir Vovk. The sources of Kolmogorov's *Grundbegriffe.* Working Paper # 4, www.probabilityandfinance.com, 2003.

[12] C. A. B. Smith. Personal probabilities and statistical analysis. *Journal of the Royal Statistical Society, Series A*, 128:469–499, 1965.

[13] Vladimir Vovk and Glenn Shafer. The game-theoretic capital asset pricing model. Working Paper # 1, www.probabilityandfinance.com, 2002.

[14] Peter Walley. *Statistical Reasoning with Imprecise Probabilities.* Chapman and Hall, London, 1991.

[15] Peter M. Williams. Indeterminate probabilities. In M. Przełęcki, K. Szaniawski, and R. Wojcícki, editors, *Formal Methods in the Methodology of Empirical Sciences*, pages 229–246. Ossolineum & Reidel, 1976.