Why Epistemology Can't be Operationalized

Timothy Williamson

University of Oxford

Timothy.williamson@philosophy.ox.ac.uk

ABSTRACT: Operational epistemology is, to a first approximation, the attempt to provide cognitive rules such that one is in principle always in a position to know whether one is complying with them. In *Knowledge and its Limits*, I argue that the only such rules are trivial ones. In this paper, I generalize the argument in several ways to more thoroughly probabilistic settings, in order to show that it does not merely demonstrate some oddity of the folk epistemological conception of knowledge. Some of the generalizations involve a formal semantic framework for treating epistemic probabilities of epistemic probabilities and expectations of expectations. The upshot is that operational epistemology cannot work, and that knowledge-based epistemology has the right characteristics to avoid its problems.

I

As advice to the enquirer, ‘Believe what is true’ is of notoriously little use. It provides no method for believing what it is true. We might try to make the point precise by defining a *method* as a set of rules such that one is, at least in principle, always in a position to know whether one is complying with them. Some are tempted to suppose that a good project for epistemology, if not the only one, would be to provide enquirers with methods in that strict sense. Such a method need not guarantee true conclusions; it might just make them more likely. Call the attempt to provide such methods *operational epistemology*.

The operational epistemologist may allow some value to non-operational epistemology, in the third-personal assessment of enquiry, which may take into account facts unavailable to the enquirers under assessment. But, it may be suggested, we also need to do epistemology in a first-personal way, in which epistemologists think of the enquirers as themselves. More precisely, it should provide enquirers with guidance which they can actually use, in whatever situation they find themselves. Rationality might be conceived as requiring conformity with the norms of operational epistemology, but not with the norms of non-operational epistemology (such as true belief). Then the thought is that the demands of rationality should be transparent to the agent: ignorance and error as such may be failures, but not failures of rationality.

Consider the injunction to proportion your confidence in a hypothesis to its probability, conditional on what you know. That injunction has no place in operational epistemology, for one is not always in a position to know how much one knows. In a situation indiscriminable from one in which one knows something, one may fail to know

without being in a position to know that one fails to know. One may also know without being in a position to know that one knows. Operational epistemologists need to work with a notion of evidence on which one is always in a position to know how much is part of one's evidence and how much is not, if they are to formulate the injunction that one should proportion one's confidence in a hypothesis to its probability, conditional on one's evidence.

Call a condition *luminous* if and only if, necessarily, whenever it obtains one is in a position to know that it obtains (given that one is a conscious agent with the relevant concepts). That one is complying with a given method in the sense of operational epistemology is a luminous condition. Thus operational epistemology and the corresponding conception of rationality require the existence of many non-trivial luminous conditions.

In *Knowledge and its Limits* (Williamson 2000: 93-113) I argued that virtually the only luminous conditions are trivial ones, which obtain always or never. Let C be a non-trivial luminous condition, which obtains sometimes but not always. Imagine a process of very gradual change from a case in which C clearly obtains to a case in which C clearly fails to obtain; throughout, the agent is considering whether C obtains. Let t_0, …, t_n be a succession of times at very brief intervals from the beginning of the process to the end. Informally, the difference between the agent's situation at t_i and at t_{i+1} is too small to be recognized by the agent; it is below the agent's threshold of discrimination. Suppose that C obtains at t_i. Since C is luminous, C obtains at t_i only if the agent is then in a position to know that C obtains. But when the agent is in a position to know that a condition obtains and is considering whether it obtains, the agent does know that the condition obtains.

Thus the agent knows at t_i that C obtains. I argued that the agent knows at t_i that C obtains only if C does obtain (whether or not the agent knows it) at t_{i+1}. The argument worked by consideration of the way in which knowledge implies a sort of reliably true belief; details are omitted here. Thus, given the premise that C obtains at t_i, one can deduce that C obtains at t_{i+1}. Since i was arbitrary, the inference works whenever $0 \leq i < n$. But, by hypothesis, C clearly obtains at t_0. Thus, iterating the inference n times, one finally reaches the conclusion that C obtains at t_n. But that is absurd; by hypothesis, C clearly fails to obtain at t_n. Given that the other features of the example can all be realized, as they can in the relevant cases, what must be given up is the assumption that C is luminous. By contraposition, its luminosity forces its triviality. The triviality of luminous conditions undermines operational epistemology. The operational epistemologist's conceptions of rationality and evidence cannot be satisfied.

Evidently, the anti-luminosity argument rests considerable weight on the reliability considerations underlying the auxiliary conditional premise that the agent knows at t_i that C obtains only if C obtains at t_{i+1}. That premise has indeed been challenged (such as Brueckner and Fiocco 2002, Neta and Rohrbaugh 2004 and Conee 2005). The specifics of those challenges will not be discussed here. However, some general remarks may clarify the position, and enable readers to work out for themselves how detailed responses to those critics would go.

First, the auxiliary conditional premise is not intended as a lawlike generalization; it is simply a description of a specific hypothetical process, justified by construction of the example. That the agent knows at t_i is not claimed to entail that C obtains at t_{i+1}

independently of the continuation of the gradual process at the later time. In particular, it depends on the fact that, by hypothesis, the agent remains alive at the later time.

Second, it is crucial to the example that, again by hypothesis, no sudden drastic change occurs during the process in the basis on which the agent believes that C obtains, that is, in the way in which the belief is formed or maintained, even if the agent is completely unaware of the change. The basis is not confined to phenomena of which the agent is conscious. For instance, one may know and believe on a highly reliable basis at one time, even though, a moment later, without one's realizing, it a mad scientist interferes with one's brain to make the basis of one's subsequent beliefs highly unreliable.

Third, even granted that the basis of belief is inaccessible to consciousness, reliability may not be determined by local properties of the basis. For instance, if someone continues over some time to believe that Lincoln is President on the basis of automatic updating, without receiving further confirmation, the reliability of the basis may depend on whether Lincoln is about to be assassinated.

The main aim of the present paper is to develop analogues of the anti-luminosity considerations for probabilistic concepts in place of the concept of knowledge. Such analogues generalize the original argument and act as a test of its soundness and significance. For someone might suspect the anti-luminosity argument of exploiting peculiarities of the concept of knowledge, either to conceal a subtle fallacy or to derive a result that, although correct, merely reflects a peculiarity of western European culture or of human folk epistemology more generally. Such a critic may hold that, in order to cut cognition at its joints, we should theorize primarily not in terms of knowledge but of

probability, a more sophisticated and graded concept. However, if we can still establish analogues of the anti-luminosity conclusions even in the new terms, then the original result is shown to be robust after all. More specifically, such generalizations of the original argument show that operational epistemologists cannot survive by formulating their test for an operational method in probabilistic terms.

The relevant concept of probability is *epistemic* or *evidential probability*, probability on one's total evidence. Now in *Knowledge and its Limits* (2000: 209-237) I argued that such evidential probability itself depends on knowledge, because one's total evidence just is the total content of one's knowledge; that is the equation E = K. Obviously, to rely on that equation here would violate the point of the exercise. The equation E = K will therefore not be used as a premise in what follows. Probabilistic claims will be assessed in probabilistic terms, without reference to questions of knowledge as such. However, in the upshot, similarities between the anti-luminosity conclusions for knowledge and for probability count against the idea that our epistemological thinking is badly distorted by our focus on knowledge rather than on probability.

II

Non-trivial conditions have been thought to be luminous because they have been regarded as epistemically perfectly accessible to the agent, completely open to view. When such a condition obtains, there should not even be a small non-zero probability that

it does not obtain. Thus the natural analogue of luminosity in probabilistic terms may be defined thus:

(LP) Condition C is *luminous in probability 1* if and only if in every case in which C obtains, the probability that C obtains is 1.

Such a condition obtains with certainty whenever it obtains at all. 'Probability' in (LP) is to be understood as evidential probability, which is not assumed to be understood specifically in terms of knowledge. In this sense, probability is not objective chance; a natural law may have an objective chance of 1 of obtaining even though nobody has the slightest reason to think that it obtains.

We shall not attribute any purely computational limitations to the agents in question. Although extant forms of would-be operational epistemology tend to make very severe demands on agents' computational capacities, in failing to meet such demands one may be considered to fall short of perfect rationality; such a verdict would be quite inappropriate in the cases to be considered here.

As before, imagine a process of very gradual change from a case in which C clearly obtains to a case in which C clearly fails to obtain, and let t_0, …, t_n be a succession of times at very brief intervals from the beginning of the process to the end. Informally, the difference between the agent's situation at t_i and at t_{i+1} is below the agent's threshold of discrimination. In probabilistic terms, it seems natural to say that it is not certain for the agent at t_i that the situation is not the one that in fact obtains at t_{i+1}, although the agent might not describe the situation like that. It is not certain exactly where one is in the

gradual process. Thus if something obtains at the later time, the epistemic probability at the slightly earlier time that it obtains (at that earlier time) is non-zero.[1] Consequently, if something fails to obtain at the later time, the probability at the earlier time that it fails to obtain (at the earlier time) is non-zero. So if the probability at the earlier time that the condition obtains is one, and the probability then that it fails to obtain is therefore zero, by contraposition the condition obtains at the later time. Hence, for a condition C:

(1_i) If at t_i the probability that C obtains is 1, then at t_{i+1} C obtains.

Once again, it must be emphasized that (1_i) is merely a definition of a hypothetical process; it does not pretend to be a law of any sort.

Now suppose that C is luminous in probability 1. Then, by (LP):

(2_i) If at t_i C obtains, then at t_i the probability that C obtains is 1.

Suppose also:

(3_i) At t_i C obtains.

By modus ponens on (2_i) and (3_i):

(4_i) At t_i the probability that C obtains is 1.

By modus ponens on (1_i) and (4_i):

(3_{i+1}) At t_{i+1} C obtains.

By construction of the example:

(3_0) At t_0 C obtains.

By repeating the argument from (3_i) to (3_{i+1}) n times, for ascending values of i from 0 to n-1, starting from (3_0):

(3_n) At t_n C obtains.

But (3_n) is false by construction of the example. Thus the premises (1_0), …, (1_{n-1}), (2_0), …, (2_{n-1}) and (3_0) together entail a false conclusion. Consequently, not all of those premises are true. But it has been argued that (1_0), …, (1_{n-1}) and (3_0) are all true. Thus not all of (2_0), …, (2_{n-1}) are true. But all of the latter follow from C's being luminous in probability 1. Thus C is not luminous in probability 1.

In its overall structure, the argument is isomorphic to the original anti-luminosity argument. It provides an extremely general template for arguing that a given condition is not luminous in probability 1. Of course, it is not completely universal, since it does not apply to conditions that always or never obtain: it assumes the possibility of a very gradual process of change that starts with cases in which C obtains and ends with cases in

which C does not obtain. Evidently, one that starts with cases in which C does not obtain and ends with cases in which C obtains would do just as well.

Although the argument bears an obvious structural similarity to a sorites paradox, it does not trade on vagueness. In particular, it does not exploit any vagueness in the concept of probability or in the specification of condition C.

The epistemic reading of 'probability' is crucial to the case for the premises (1_0), …, (1_{n-1}). The case concerns the agent's epistemic limitations, not objective chances. If natural laws and present circumstances determine that the condition obtains at one time, they may still determine that it does not obtain at a very slightly later time. But epistemic probability should also not be confused with pure subjective probability, the agent's credence or degree of belief. For all that has been said, the agent might be irrationally quite certain throughout the process that C obtains, in which case (1_{n-1}) would be false if 'probability' were read as degree of belief. The justification for (1_0), …, (1_{n-1}) is that, during the gradual process, if C fails to obtain at one time then it was already not certain on the agent's evidence very shortly before that C obtained then, however subjectively certain the agent was that it obtained. This sort of epistemic probability is far more relevant than pure subjective probability to any normative epistemology or philosophy of science that attempts to get beyond mere criteria of internal coherence.[2]

III

Could a philosopher react to the argument of sections II by lowering the demand from luminosity in probability 1 to luminosity in probability x, for some real number x strictly between 0.5 and 1? For example, a condition C is *luminous in probability 0.9* if and only if in every case in which C obtains, the probability that C obtains is at least 0.9. To argue in parallel fashion that C is not luminous in probability x, one would need in place of premise (1_i) the stronger claim that if at t_i the probability that C obtains is at least x, then at t_{i+1} C obtains. That stronger claim might be hard to support. Formally, in some models, at each of the times t_0, …, t_n the probability that one is at that time is 0.9, with the remaining 0.1 of probability distributed equally amongst the immediately neighbouring times: such models make every condition whatsoever luminous in probability 0.9 while acknowledging that it is not certain exactly where one is in the gradual process.

As already hinted, such watered-down probabilistic notions of luminosity do not satisfy the motivations for postulating non-trivial luminous conditions in the first place. If it is probable but not certain that my evidence is what I think it is, then my failure to proportion my beliefs to my evidence might reflect my incomplete information about my evidence, and so be difficult to classify as irrationality on the operational epistemologist's conception. If it can be probable without being certain for one that one seems to oneself to see a red patch, then the rule 'Say "I seem to myself to see a red patch now!" when and only when one seems to oneself to see a red patch' cannot be part of a method in the operational epistemologist's sense. Luminosity only in probabilities less than 1 has many

of the same philosophical consequences as anti-luminosity. Let us therefore return to the consideration of luminosity in probability 1.

IV

The argument of section II has a problematic feature. In explaining it, we can start by using times as convenient proxies for the relevant aspects of the situations at those times. Then the argument in effect assumes that when one is at t_i there is a non-zero epistemic probability c that one is at t_{i+1} instead. That assumption may look plausible if time is discrete, so that each moment has an immediate predecessor and an immediate successor. But if time is dense then there are infinitely many times between t_i and t_{i+1}. Presumably, for each of those intermediate times t, when one is at t_i, the epistemic probability that one is at t is at least as great as the epistemic probability that one is at t_{i+1}. Thus at t_i infinitely many disjoint events (incompatible propositions) have probability at least c, which is impossible because the probabilities of disjoint events must sum to at most 1. The natural assumption is rather that time is continuous and for any two times t and t^*, the probability when one is at t^* that one is at t is 0. Thus premise (1_i) is too strong. If condition C obtains at exactly t_{i+1} but at no other time, then the probability at any time (including both t_i and t_{i+1}) that C does not obtain may still be 1. In continuous probability spaces, probability 1 does not even entail truth.

Of course, what is typically at issue is the agent's information, not about the time itself, but about some aspect of the situation that changes over time, such as the agent's

mental state. Hence what really matters is the discreteness or continuity of that aspect rather than of time itself. If time is continuous, some changing aspect of the situation may still be discrete: for example, the number of lights that are on. If time is discrete, some changing aspect of the situation may still take values from a continuous space of possibilities; even if only finitely many of those possibilities are actually realized in a period of finite length, infinitely many other possibilities may be epistemically at least as probable from the agent's perspective. Since many philosophically significant aspects of the agent's situation at least appear to admit of continuous variation, and that appearance cannot simply be dismissed as illusory, we must take the problem of continuous variation seriously. Much of the discussion will continue to be phrased in terms of the agent's information about the time, but it should not be forgotten that this is merely a convenient expository device.

One could try to fix the problem of continuous variation by permitting infinitesimal non-zero probabilities, since infinitely many non-zero infinitesimals can have a finite sum. However, such a reliance on non-standard probability theory would slightly weaken the argument by increasing its burden of assumptions. Let us therefore stick to standard real-valued probabilities and the assumptions that time is ordered like the real numbers and that for any two times t and t^*, the probability when one is at t^* that one is at t is 0.

Another attempted fix would treat the probabilities in question as conditional on the supposition that one is exactly at one of the times $t_0, \ldots, t_n$, while still allowing that time itself is continuous. However, if the (unconditional) probability that one is at a given t_i is 0, then the (unconditional) probability that one is at one of $t_0, \ldots, t_n$ is also 0; we

cannot in general assume that probabilities conditional on a supposition of (unconditional) probability 0 are well-defined. The technical problem is that the probability of p conditional on q $P(p|q)$ is usually defined as the ratio of unconditional probabilities $P(p\&q)/P(q)$, so that probabilities conditional on a supposition of (unconditional) probability 0 correspond to dividing by 0. One might instead take the conditional probability $P(p|q)$ as primitive and undefined, but that too would slightly weaken the argument by increasing its burden of assumptions.

A more natural approach works with arbitrary brief open intervals of time. Henceforth all talk of times and intervals is to be understood as restricted to times and intervals within the closed interval $[t_0, t_n]$, during which the gradual process takes place. The idea is that some short but non-zero length of time ε is such that at any time t^* there is a non-zero probability that one is in any given interval of non-zero duration within ε of t^*: at t^*, one has no certainty as to where one is in that interval. Thus if t and t^{**} are two times such that $t^*-\varepsilon \leq t < t^{**} \leq t^*+\varepsilon$, then the probability at t^* that one is in the open interval (t, t^{**}) is non-zero. In effect, the quantity ε corresponds to the length of time from t_i to t_{i+1} in the discrete case, which was assumed independent of i (if the duration varied with i, one could simply choose a duration less than the minimum of those lengths for $i = 0, \ldots, n-1$). Similarly, it was assumed that ε can be chosen small enough to be independent of t^*. This assumption is entirely plausible in the sorts of example with which we are concerned. Over the whole gradual process there is a finite limit to the agent's powers of discrimination; they do not become arbitrarily fine within the interval.

Now suppose that a condition C* obtains throughout the interval (t, t^{**}) $(t < t^{**})$, and that t^* is in the longer interval $(t-\varepsilon, t^{**}+\varepsilon)$. Hence $t < t^*+\varepsilon$ and $t^*-\varepsilon < t^{**}$, so $\max\{t,$

t^*-ε} < min{t^{**}, t^*+ε}. Thus C* obtains throughout the interval of non-zero duration (max{t, t^*-ε}, min{t^{**}, t^*+ε}), which is within the interval (t^*-ε, t^*+ε). Consequently, by what was said in the previous paragraph, the probability at t^* that C* obtains should be non-zero; as before, knowledge of the time was being used merely as a convenient proxy for knowledge of the conditions obtaining at the time. Now let C* be the negation of condition C. We have in effect derived this constraint, similar to the contraposed version of (1_i):

(1) If C obtains at no time in a nonempty interval (t, t^{**}), then at no time in the interval (t-ε, t^{**}+ε) is the probability that C obtains 1.

Now suppose again that C is luminous in probability 1. Consequently, by (LP), we have this analogue of the contraposed version of (2_i):

(2) If at no time in the interval (t-ε, t^{**}+ε) is the probability that C obtains 1, then C obtains at no time in the interval (t-ε, t^{**}+ε).

Combining (1) and (2) yields:

(5) If C obtains at no time in a nonempty interval (t, t^{**}), then C obtains at no time in the interval (t-ε, t^{**}+ε).

But in the examples with which we are concerned, C does not just clearly fail to obtain at t_n; that time was chosen to be well within the period in which C fails to obtain, so that for some small but non-zero duration δ:

(6_0) C obtains at no time in the interval (t-δ, t).

Applying (5) i times to (6_0) yields this:

(6_i) C obtains at no time in the interval (t-δ-iε, t+iε).

By taking i large enough, we therefore derive:

(7) C obtains at no time in the interval [t_0, t_n].

But (7) is absurd, for by construction of the example C clearly obtains at t_0. Just as in section II, this constitutes a reductio ad absurdum of (LP); condition C is not luminous in probability 1. Although the argument differs in detail from that of section II, its spirit is recognizably the same.

Since the new argument requires an example in which the condition at issue fails to obtain throughout a small open interval, it is inapplicable to a slightly wider range of conditions than the original anti-luminosity argument was. Consider a condition for which between any two moments at which it obtains there is a moment at which it does not obtain and between any two moments at which it does not obtain there is a moment at

which it does obtain (the rational and irrational numbers are so related). Such a condition satisfies (1) vacuously. Yet it need not violate (LP), for on some distributions at every moment it has probability 1 of obtaining. But such curiosities are irrelevant to the original motivations for postulating non-trivial luminous conditions. The conditions at issue, such as seeming to see a red patch, surely can fail to obtain throughout an extended period.

The upshot of this section is that although the continuity of time complicates the form of the argument that a given condition is not luminous in probability 1, it does not undermine the argument in its philosophically most significant applications.

V

The continuity of time makes it even less promising to replace luminosity in probability 1 by luminosity in a probability strictly between 0.5 and 1 than it was in the discrete case (discussed in section III). For the simplest assumption is that at any time t^* far from the beginning and from the end of the process the epistemic probability that one is at a time before t^* is the same as the epistemic probability that one is at a time after t^*, while the epistemic probability that one is exactly at t^* itself is 0; although the assumption need not hold in all examples, it should at least hold in some. In those examples, the probability at t^* that one is at a time before t^* and the probability that one is at a time after t^* are both 0.5. Suppose that, during the gradual process, a long period throughout which the condition C obtains is immediately followed by a long period throughout which C fails to obtain. If t^* is a time in the former period, but close to its end, then the probability at t^*

that C obtains is in effect the probability that one is either at a time before t^* or at a time in the brief period after t^* while C still obtains, which is 0.5 plus the probability that one is in the brief period. By choosing t^* closer and closer to 0.5, we should be able to make the probability that one is in the brief period closer and closer to 0. Consequently, for any real number x greater than 0.5, there should be a time at which C obtains but the probability that C obtains is less than x, in other words, a counterexample to the claim that C is luminous in probability x.

Under the foregoing assumptions, the best to be hoped for is that whenever C obtains, the probability that C obtains is greater than 0.5. That is slightly stronger than luminosity in probability 0.5, since it has 'greater than' in place of 'at least'. We can model the stronger claim by supposing that at each time t all the probability is uniformly distributed over the interval $(\max\{t_0, t-\varepsilon\}, \min\{t_n, t+\varepsilon\})$, and that (for some particular condition C) whenever C obtains at a time t it obtains throughout an open interval containing t of length greater than ε. But even the stronger claim is strikingly weak in relation to the original philosophical motivations for postulating interesting luminous conditions. To say that when something is part of one's evidence it is more probable than not for one that it is part of one's evidence falls far short of what would be required for the rule 'Proportion your confidence in a hypothesis to its probability on your evidence' to count as fully operational in the usual sense.

VI

An alternative fallback from luminosity in probability 1 uses the mathematical notion of the *expectation* of a real-valued random variable, its mean value weighted by probability. A 'random variable' in this technical sense need not be perfectly random; any function from worlds or states to real numbers (such as the length of a given object in centimetres) constitutes a random variable. Let the agent's probability distribution at time *t* be P. If the random variable X takes values from the finite set I, then its expectation E(X) at time *t* is defined by the equation:

(E) $\quad E(X) = \sum_{x \varepsilon I} x P(X = x)$

If X takes values in an infinite set, then the finite sum in (E) will need to be replaced by an infinite sum or integral. For present purposes we may omit mathematical details and harmlessly assume that the expectations with which we are concerned all have well-defined values. Now even if the agent cannot be certain what the value of X is, the expected value of X may still equal its actual value. Moreover, the equality may be systematic. If the actual value of a variable is always its expected value for the agent, then the agent seems to have a special probabilistic sort of privileged access to that variable, potentially relevant to the theory of rationality.

Here is a simple example, which does not pretend to be realistic. Suppose that times can be regarded as the positive and negative integers. The variable T is the time. At each time *t*, it is equally likely that the time is any one of *t*-1, *t* and *t*+1, but certain that it is one of them: $P(T = t\text{-}1) = P(T = t) = P(T = t\text{+}1) = 1/3$. Consequently, at *t*, $E(T) = (t\text{-}1)/3 + t/3 + (t\text{+}1)/3 = t$. Thus the expected value of the time is always its actual value, even though

one can never be certain what its actual value is. A similar result holds if time is ordered like the real numbers, provided that at each moment probability is symmetrically distributed over past and future.

A tempting line of thought might seduce one into thinking that such a case cannot arise, at least if the agent has access to permanent features of the epistemic situation. For if one knows at *t* that the expected time is *t*, and that the expected time is always the actual time, one can deduce that the time is *t*, contrary to the original stipulation that at *t* the epistemic probability that the time is *t* is only 1/3.

The crucial fallacy in the tempting line of thought is the assumption that the agent is always in a position to know what the expectation of a variable is. By the model, at *t* there is a 1/3 probability that the expectation of the time is *t*-1 and a 1/3 probability that it is *t*+1 (although the expectation of the expectation of the time is still *t*). This uncertainty about the values of expectations derives from uncertainty about the values of probabilities: for example, at *t* there is a 2/3 probability that there is a 1/3 probability that the time is *t*+1 (since at both *t* and *t*+1 there is a 1/3 probability that the time is *t*+1), but there is also at *t* a 1/3 probability that there is probability 0 that the time is *t*+1 (since at *t*-1 there is probability 0 that the time is *t*+1).

One might wonder what use expectations are to an agent who cannot be sure what their values are. But that is just to hanker again after luminosity in probability 1 (in this case, of the expectations), which we have already seen to be unobtainable in the cases that matter. If we really do face uncertainty all the way down, a prospect with which it is by no means easy to be comfortable, then variables whose uncertain actual values always coincide with their equally uncertain expected values might still have a privileged role to

play in a sober form of operational epistemology. At least they do not seem quite as badly off as variables whose actual values differ from their expected values. More specifically, one cannot operationalize such a rule any further by replacing such a random variable in the rule with its expectation, for the replacement makes no difference.[3]

Amongst the random variables whose expectations we sometimes wish to consider are expectations and probabilities themselves. Indeed, we can regard probabilities as special cases of expectations. For let the *truth-value* of A be 1 if A is true and 0 otherwise. Then the probability of A is simply the expectation of the truth-value of A. Thus we shall be concerned with expectations of expectations and probabilities of probabilities. These higher-order notions need delicate handling. Before proceeding further, it is therefore prudent to develop a rigorous formal semantic framework for their treatment. We can then investigate the significance of the anti-luminosity considerations for variables whose actual and expected values coincide.

VII

We first construct a formal language for the expression of higher-order probabilities and expectations. For simplicity, the language does not include quantifiers. Adding them would involve deciding on a policy for handling the tricky problem of quantifying into epistemic contexts such as probability operators create. However, the language does include predicates and singular terms; the latter will all be understood as denoting real numbers.

We define the formulas and terms of the language by a simultaneous recursion:

There are countably many atomic variables X, Y, Z, … (informally, denoting real numbers; they correspond to random variables).

For each rational number *c* there is an atomic constant [*c*] (informally, denoting *c*).

If T and U are terms then T+U is a term (+ is informally read as 'plus').

If A is a formula then ¬A is a formula (¬ is informally read as 'it is not the case that').

If A and B are formulas then A & B is a formula (& is informally read as 'and').

If A is a formula then V(A) is a term (V is informally read as 'the truth-value of').

If T and U are terms then T≤U is a formula (≤ is informally read as 'is less than or equal to').

If T is a term then E(T) is a term (E is informally read as 'the expectation of').

Other truth-functors can be introduced as meta-linguistic abbreviations in the usual way. Moreover, 'T=U' abbreviates T≤U & U≤T and 'T<U' abbreviates T≤U & ¬U≤T

(mathematical symbols will be used in both object-language and meta-language since they are being used with the same sense). The expression 'P(A)' for the probability of A abbreviates E(V(A)). The inclusion of infinitely many atomic constants is for mathematical convenience only. In particular applications only finitely many such constants are needed.

We must now give a model-theoretic semantics for the language. A model is a triple <W, Prob, F>, where W is a nonempty set, Prob is a function from members of W to probability distributions over W and F is a function from atomic terms to functions from members of W to real numbers. Informally, W is the set of states or worlds; for a world $w \varepsilon W$, Prob(w) is the epistemic probability distribution for the agent at w; if T is an atomic term, F(T)(w) is the value of the random variable T at w. We define the truth-value at w of A, val(w, A), and the denotation at w of T, den(w, T), for all worlds w, formulas A and terms T by another simultaneous recursion:

If T is an atomic variable, den(w, T) = F(T)(w).

If c is a rational number, den(w, [c]) = c.

If T and U are terms, den(w, T+U) = den(w, T) + den(w, U).

If A is a formula, val(w, ¬A) = 1-val(w, A).

If A and B are formulas, val(w, A & B) = min{val(w, A), val(w, B)}.

If A is a formula, den(w, V(A)) = val(w, A).

If T and U are terms, val(w, T≤U) = 1 if den(w, T) ≤ den(w, U) and 0 otherwise.

If T is a term, den(w, E(T)) is the expectation with respect to the probability distribution Prob(w) of the random variable whose value at each world xεW is den(x, T).

To illustrate the last clause, if W is finite, then:

$$\text{den}(w, \text{E(T)}) = \sum_{x \varepsilon W} \text{den}(x, \text{T})\text{Prob}(w)(\{x\})$$

When W is infinite, an analogous infinite sum or integral applies. We count <W, Prob, F> as a model only if the expectation den(w, E(T)) is well-defined for every wεW and term T of the language.

This semantics determines truth-conditions for statements involving arbitrary iterations of the probability and expectation operators. It was the tacit background to much of the preceding discussion, and can be used to make it precise. One striking effect of the semantics is that the expectation of the expectation of a variable need not be its expectation; in other words, some expectations in some models fall outside the class of random variables whose actual and expected values always coincide. Consider the following model <W, Prob, F>. It has three worlds; W = {*u, v, w*}. Informally, think of

the worlds close to *u* as just *u* itself and *v*, of the worlds close to *w* as just *v* and *w* itself, and of all worlds as close to *v*. Then Prob maps each world to the probability distribution that makes all the worlds close to that one equiprobable, and assigns probability zero to any other world. Thus Prob(*u*)({*u*}) = Prob(*u*)({*v*}) = ½ and Prob(*u*)({*w*}) = 0; Prob(*v*)({*u*}) = Prob(*v*)({*v*}) = Prob(*v*)({*w*}) = 1/3; Prob(*w*)({*u*}) = 0 and Prob(*w*)({*v*}) = Prob(*w*)({*w*}) = ½. Treat X as the random variable with these values: den(*u*, X) = F(X)(*u*) = 8; den(*v*, X) = F(X)(*v*) = 4; den(*w*, X) = F(X)(*w*) = 0. It is now elementary to calculate that den(*u*, E(X)) = 6, den(*v*, E(X)) = 4 and den(*w*, E(X)) = 2. Iterating the process, we calculate that den(*u*, E(E(X))) = 5, den(*v*, E(E(X))) = 4 and den(*w*, E(E(X))) = 3. Thus the expectation of the expectation of X differs from the expectation of X at the worlds *u* and *w*, although not at *v*.

A paradox may seem to threaten. For is not one of the most elementary facts that students learn about calculating expectations the equation 'E(E(X)) = E(X)'? But the expression 'E(E(X))' conceals an ambiguity. For since 'E(X)' is in effect a definite description, a question arises about its scope in the context 'E(E(X))'. The standard textbook treatment assigns 'E(X)' wide scope in 'E(E(X))'; informally, the expectation of the expectation of X is simply the expectation of that constant whose value is always *c*, where *c* is in fact the actual expectation of X. This reading is of course perfectly intelligible and legitimate; it makes the equation 'E(E(X)) = E(X)' trivially correct and does not really raise any issue of higher-order expectations or probabilities. The value of 'E(E(X))' on the wide scope reading can be calculated on the basis of a single probability distribution. But equally intelligible and legitimate is a reading on which 'E(X)' has narrow scope in 'E(E(X))'. The semantics above generates the latter reading and is quite

natural. It is this reading which we need in order to raise genuine questions about higher-order expectations and probabilities, and it does not make the equation 'E(E(X)) = E(X)' trivially correct. The value of 'E(E(X))' on the narrow scope reading depends on a probability distribution over probability distributions.

We can see the difference between the two readings in practice by considering the defining equation for the *variance* of a random variable, 'Var(X) = E((X−E(X))^2)'. Var(X) is supposed to measure how much X is expected to deviate from its mean. Consider the example in section VI of a random variable whose actual and expected values always coincide. Since the equation 'E(X) = X' is always certain to be true, the value of (X−E(X))^2 is always certain to be 0, so the value of 'E((X−E(X))^2)' on the narrow scope reading is always 0 too. But of course on the wide scope reading, the value of 'E((X−E(X))^2)' at any time *t*, where E(X) = *t*, is calculated in the usual way as $((t-1-t)^2 + (t-t)^2 + (t+1-t)^2)/3 = 2/3$. Both readings give genuine information, but they given different information. The wide scope reading tells us that for any number *c*, if *c* is in fact the expected value of X then X is expected to deviate from *c* by a certain amount. The narrow scope reading tells us that it is certain that X will nor deviate from its expected value.

Having clarified these matters, let us apply the semantics above to determine whether variables whose actual and expected values always coincide constitute a useful fallback from probabilistic luminosity in the previous senses.

VIII

Let us start with the simple case in which the set W of worlds is finite. The anti-luminosity considerations trade on the possibility of constructing sorites series between radically different cases. Suppose that such series always exist in W, in this sense:

SORITES For any $w, x \in W$, there are $w_0, \ldots, w_n \in W$ such that $w = w_0$, $x = w_n$ and for $0 \leq i, j \leq n$, if $|i-j| \leq 1$ then $\text{Prob}(w_i)(\{w_j\}) > 0$.

The idea is that at any world it is uncertain whether one is at that world or at one of its immediate neighbours in the series. In particular, for $0 \leq i < n$, $\text{Prob}(w_i)(\{w_{i+1}\}) > 0$ by SORITES. Very roughly, one can get from any world to any world via a chain of worlds each of which is imperfectly discriminable from its predecessor. In this sense, one world is imperfectly discriminable from another if and only if it is not certain for the agent at the latter world that the former does not obtain.

In this setting we can now reconstruct the argument against luminosity. Suppose that we are interested in the value of a term T, and that its expected and actual values always coincide. That is:

(#) For all $w \in W$, $\text{den}(w, T) = \sum_{x \in W} \text{den}(x, T)\text{Prob}(w)(\{x\})$

Since W is finite, $\text{den}(w, T)$ attains a maximum value max(T) at some world w. Let MAX(T) be the set of worlds at which T attains this maximum, $\{x \in W: \text{den}(x, T) =$

max(T)}. Suppose that $y \varepsilon \text{MAX(T)}$, and that the world z is imperfectly discriminable from y: $\text{Prob}(y)(\{z\}) > 0$. Consequently, by (#):

$$\begin{aligned} \max(\text{T}) &= \text{den}(y, \text{T}) = \textstyle\sum_{x \varepsilon \text{W}} \text{den}(x, \text{T})\text{Prob}(y)(\{x\}) \\ &\leq \max(\text{T})(1\text{-Prob}(y)(\{z\}) + \text{den}(z, \text{T})\text{Prob}(y)(\{z\}) \\ &= \max(\text{T}) - (\max(\text{T}) - \text{den}(z, \text{T}))\text{Prob}(y)(\{\text{z}\}). \end{aligned}$$

Thus $(\max(\text{T}) - \text{den}(z, \text{T}))\text{Prob}(y)(\{\text{z}\}) \leq 0$. But $\text{den}(z, \text{T}) \leq \max(\text{T})$ and $\text{Prob}(y)(\{z\}) > 0$, so $\text{den}(z, \text{T}) = \max(\text{T})$, so $z \varepsilon \text{MAX(T)}$. Thus any world imperfectly discriminable from a world in MAX(T) is itself in MAX(T). Consequently, any chain of imperfect discriminability that starts in MAX(T) remains wholly in MAX(T). Hence, by SORITES, any world in W can be reached from a world in MAX(T) by a chain of imperfect discriminability. Therefore, every world in W is in MAX(T). So T attains its maximum at every world in W: in other words, T is constant.

What we have proved is that, in the finite case, SORITES implies that a random variable whose expected and actual values always coincide is trivial, in the sense that its value cannot vary across worlds. But that rules out the variables in which we are most interested, since they do differ in value between worlds (or states) that are joined by a series of imperfectly discriminable intermediaries. In particular, the agent's evidence and mental state vary over such sorites series. Consequently, at least when the state space is finite, it is useless to postulate an epistemically privileged set of variables whose actual and expected values always coincide. Those variables would exclude everything that can be learnt from new experiences that develop gradually out of old ones.

The same argument sometimes applies even when the set of worlds is countably infinite, although the condition that the term T attains a maximum value at some world is

no longer automatically satisfied. The probability distributions required for SORITES may also be less natural in that case, since one cannot give the same non-zero probability to infinitely many worlds. If the set of worlds is uncountably infinite, SORITES must fail, because at most countably many worlds are imperfectly discriminable in the relevant probabilistic sense from any given world, so at most countably many worlds are linked by a finite chain of imperfect discriminability to a given world. Thus the infinite case requires separate consideration.

An easy generalization is to cases in which, although there are infinitely many worlds, the term T takes values only from a finite set. A more useful generalization is to all cases in which T attains a maximum value without certainty that it does so:

UNCMAX For some $w \varepsilon W$: for all $x \varepsilon W$, $den(x, T) \leq den(w, T)$, but
$Prob(w)(\{x: den(x, T) < den(w, T)\}) > 0$.

By the second conjunct of UNCMAX, $Prob(w)(\{x: den(x, T) + 1/n \leq den(w, T)\}) > 0$ for some natural number n.[4] Let $Prob(w)(\{x: den(x, T) + 1/n \leq den(w, T)\})$ be p; thus $p > 0$. Then

$$den(w, E(T)) \leq den(w, T)(1-p) + (den(w, T) - 1/n)p = den(w, T) - p/n < den(w, T).$$

Thus it is false at w that the actual and expected values of T coincide. Similarly, they fail to coincide when T attains its minimum without certainty that it does so.

A central application of the preceding result concerns cases in which T is itself a probability, so that E(T) is in effect the expectation of an expectation. If T attains the value 1 at some world, that is of course its maximum. Thus if an event can be certain

without being certainly certain, the actual and expected values of the probability of that event do not always coincide.[5] One might suppose certainty to imply certainty of certainty, but that is to suppose certainty to be luminous in probability 1, which by earlier results we must expect it not to be unless the standard for certainty is set almost impossibly high. Of course, some philosophers will retort that certainty is an (almost) impossibly high standard for empirical beliefs. If some event cannot attain probability 1, but can attain any probability less than 1, then the preceding argument does not apply to it. However, an event might also attain a maximum probability less than 1. For example, let *w* and *w** be similar but not perfectly similar worlds, and consider the probability of the condition C that one is at a world more similar to *w* than *w** is. Other things being equal, that probability will normally attain a maximum at *w* itself. However, at *w*, there may well be a positive probability that one is at a world slightly different from *w* at which the probability that C obtains is less than maximal; then UNCMAX is satisfied. Consequently, at *w* the expected probability that C obtains is less than the actual probability. As already noted, not even expectations have the epistemic privilege that their expected value is always their actual value.

It might still be tempting to think that even if probabilities and expectations lack a perfect epistemic privilege, they are still epistemically more privileged than what they are probabilities or expectations of. Thus one might regard the sequence of a random variable, its expectation, the expectation of its expectation, the expectation of the expectation of its expectation, … as tending in the direction of ever greater but never perfect epistemic privilege. The trouble with this idea is that the increase in epistemic privilege is typically achieved by the washing out of empirical evidence. This

phenomenon is vivid when the number of worlds is finite. Suppose that it is, and that SORITES holds: any two worlds are linked by a chain of imperfectly discriminable pairs. Let T be any random variable. Then it is provable that the sequence T, E(T), E(E(T)), … converges to the same limit *whichever world one is at*.[6] The original variable T may encode significant information about which world one is at, differentiating some worlds from others, but the process of taking iterated expectations gradually wipes out those differences. Thus the tendency in the direction of epistemic privilege is a tendency in the direction of failing to learn from experience.

As usual, the picture is more complex when the number of worlds is infinite, but it is doubtful that the extra complexities make very much difference to the overall upshot for epistemology. Under realistic assumptions, taking iterated expectations still tends to wipe out empirical differences and thereby undermine learning from experience.

IX

The more one tries to make something epistemically privileged, the less it can discriminate between different states of the world. Of course, empirical evidence must have *some* epistemic privilege, otherwise there would be no point for the epistemologist in distinguishing it from what it is evidence about. But if it has too much epistemic privilege, then it no longer discriminates between initially possible outcomes as learning requires. The notion of evidence is held in place, not altogether comfortably, by these

opposing pressures. The conception of evidence as the total content of one's knowledge (the equation E = K) is a compromise of just the required kind.

The considerations of this paper suggest that the anti-luminosity arguments of *Knowledge and its Limits* are robust. They are not symptoms of an isolated pathological quirk in the ordinary concept of knowledge, 'the epistemology of the stone age'. Rather, they can be replicated from a probabilistic starting-point, and reflect the general predicament of creatures with limited powers of discrimination in their attempt to learn from experience. As epistemologists, we need the concept of probability, but the idea that our pervasive fallibility requires that it replace the concept of knowledge is itself the product of residual infallibilistic assumptions, which attribute imaginary epistemic privileges to probability itself. We are fallible even about the best descriptions of our own fallibility. The concept of knowledge is adapted to just such a predicament. In particular, it can demarcate the evidence that conditions evidential probabilities.

By contrast, the operational epistemologist's attempt to work without such 'impure' concepts results in a futile psychologization of the concept of evidence, on which one's evidence is reduced to one's present subjective states. That concept of evidence diverges drastically from what we need to understand the objectivity of the natural sciences, yet still fails to achieve the kind of luminosity which the psychologization aimed to achieve. Like other movements driven by unrealistic idealism disguised as tough-mindedness, the operationalization of epistemology would destroy what it sets out to reform.[7]

Notes

1 One must exclude trivial indexical ways of specifying conditions such as 'The condition that the time is now' which refer to different conditions at different points in the process.

2 Some hard-line subjective Bayesians will object at this point that the notion of Jeffrey conditionalization suffices for operational epistemology. For criticism see Williamson 2000: 216-219.

3 Williamson 2000: 230-237 explores uncertain probabilities within a framework based on the concept of knowledge.

4 Since den(w, E(V(T<[den(w, T)])) and den(w, E(V(T+[1/n]≤[den(w, T)])) are well-defined by definition of a model, the probabilities in the argument and UNCMAX are also well-defined. The argument that Prob(w)({x: den(x, T) + 1/n ≤ den(w, T)}) > 0 for some n assumes that Prob is countably additive. If countably non-additive probability distributions are allowed, that derived condition should be used in UNCMAX instead.

5 In a series of papers, Dov Samet (1997, 1998, 2000) has investigated constraints in epistemic logic related to those examined here. In particular, he considers the 'averaging' axiom that the actual and expected probabilities of an event must coincide (which he

takes from Jeffrey 1992 and Skyrms 1980) and shows that in the finite case it implies certainty that if the probability of an event has a given value then it is certain that it has that value. Roughly speaking, Samet's results concern the hypothesis that the axiom is universally satisfied, whereas the present inquiry is into the hypothesis that a particular event satisfies it. Williamson (2000: 311-315) proves that on a frame for epistemic logic in which one can know without knowing that one knows or fail to know without knowing that one fails to know, unconditional prior probabilities do not always coincide with the prior expectations of posterior probabilities, where posterior probabilities at a world are the results of conditionalizing prior probabilities on the total content of what one knows at that world. That framework is more restrictive than the present one. To take the simplest case, consider a model with just two worlds, *w* and *x*. On the present framework, the event $\{w\}$ could have probability 2/3 at *w* and probability 1/3 at *x*. That cannot happen within the framework of Williamson 2000, since conditionalization is either on the whole of $\{w, x\}$, which makes no difference, or on a singleton set, in which case all probabilities go to 1 or 0: thus no event ever has a probability strictly between 0 and 1 other than its prior probability (of course, that result does not generalize to models with more than two worlds). Since that more restrictive framework is justified by the knowledge-centred approach, it cannot be assumed here.

6 Proof: One can consider the worlds as states in a Markov chain, with $\mathrm{Prob}(i)(\{j\})$ as the one-step transition probability from state *i* to state *j*. Let $p_{ij}(n)$ be the transition probability that one is at *j* after *n* steps starting from *j*. Let $\mathrm{E}^0(\mathrm{T}) = \mathrm{T}$ and $\mathrm{E}^{n+1}(\mathrm{T}) = \mathrm{E}(\mathrm{E}^n(\mathrm{T}))$. Then, by induction on *n*, for any *n* and $i \varepsilon \mathrm{W}$, $\mathrm{den}(i, \mathrm{E}^n(\mathrm{T})) = \sum_{j \varepsilon \mathrm{W}} \mathrm{den}(j, \mathrm{T}) p_{ij}(n)$.

But by SORITES the Markov chain is irreducible and aperiodic, so by a standard result about Markov chains (Grimmett and Stirzaker 2001: 232), $p_{ij}(n)$ converges to a limit that is independent of i as n goes to infinity. Consequently, den(i, $E^n(T)$) also converges to a limit that is independent of i as n goes to infinity.

7 I thank the Centre for Advanced Study at the Norwegian Academy of Science and Letters for its hospitality during the writing of much of this paper. Some of the material was presented in various earlier versions at classes in Oxford, at a conference on contextualism at the University of Stirling, and at the Jean Nicod Institute in Paris, the University of Bristol, the Graduate Center of City University New York and at the University of California at Santa Barbara; I thank the audiences for helpful comments.

Bibliography

Brueckner, A., and Fiocco, M.O. 2002. 'Williamson's Anti-Luminosity Argument'. *Philosophical Studies* **110**: 285-293.

Conee, E. 2005. 'The comforts of home'. *Philosophy and Phenomenological Research* **70**: 444-451.

Grimmett, G., and Stirzaker, D. 2001. *Probability and Random Processes*, 3rd ed. Oxford: Oxford University Press.

Jeffrey, R. 1992. *Probability and the Art of Judgement*. Cambridge: Cambridge University Press.

Neta, R., and Rohrbaugh, G. 2004. 'Luminosity and the safety of knowledge'. *Pacific Philosophical Quarterly* **85**: 396-406.

Samet, D. 1997. 'On the triviality of high-order probabilistic beliefs'. *Game Theory and Information* REPECc:wpa:wuwpga:9705001.

Samet, D. 1998. 'Iterated expectations and common priors'. *Games and Economic Behaviour* **24**: 131-141.

Samet, D. 2000. 'Quantified beliefs and believed quantities'. *Journal of Economic Theory* **95**: 169-185.

Skyrms, B. 1980. *Causal Necessity*. New Have, CT: Yale University Press.

Williamson, T. 2000. *Knowledge and its Limits*. Oxford: Oxford University Press.