Very Improbable Knowing*

(draft: comments welcome)

Timothy Williamson

University of Oxford

0. In general, the truth of a proposition entails very little about its epistemic status. It need not be known, or even knowable. It is at least not known to be false, since only truths are known, but that is not saying much. It may be true even though it is almost certain on the evidence that it is false, for evidence can be highly misleading.

Some propositions about one's present epistemic state are often regarded as exceptions to the general compatibility of truth and low epistemic status. Perhaps the most famous example is the KK principle, also sometimes called "positive introspection". It says that if one knows a truth *p*, one knows that one knows *p*. Thus the truth of the proposition that one (presently) knows *p* is supposed to entail that the (present) epistemic status of the proposition that one knows *p* is at least as good as knowledge. However, it is widely, although not universally, acknowledged that the KK principle is false, and not just for the boring reason that one can know *p* without having formed the belief that one knows *p*. One can know *p*, and believe that one knows *p*, without knowing that one knows *p*, because one is not in a strong enough epistemic position to know that one knows *p* (Williamson 2000: 114-30). ‘I know *p*’ can be lower than *p* in epistemic status. But how *much* lower in epistemic status can ‘I know *p*’ be than *p* itself?

We can grade epistemic status in terms of evidential probabilities. If one knows *p*, how improbable can it be, on one’s own present evidence, that one knows *p*? One conclusion of this paper is that the probability can sink arbitrarily close to 0. At the limit, the probability on one’s evidence of *p* can be 1 while the probability on one’s evidence that one knows *p* is 0. The difference between the probabilities can be as large as probabilistic differences can go.

One argument for such a conclusion follows a traditional way of arguing against the KK principle, by using familiar examples that support some form of externalism about knowledge (Lemmon 1967). For instance, the unconfident examinee answers questions on English history under the impression that he is merely guessing. In fact, his answers are correct, and result from lessons on it that he has completely forgotten he ever had (Radford 1966). The example can be so filled in that it is extremely improbable on the examinee’s evidence that he had any such lessons, or any other access to the relevant knowledge of English history; nevertheless, he does know the historical facts in question. That description of the example may well be right. Dialectically, however, it has the disadvantage that those of a more internalist bent may simply deny that the examinee knows. This paper develops a more systematic, structural way of arguing for the realistic possibility of knowing when it is extremely improbable on one’s evidence that one knows, a way independent of specifically externalist judgments about cases.

On the resulting view, one’s evidence can be radically misleading about one’s own present epistemic position. Since the rationality of an action depends on one’s epistemic position, one’s evidence can be radically misleading about the rationality of the various actions available to one. Such phenomena will be used to cast light on some epistemological puzzles. If our pre-theoretic assessments of particular cases neglect the possibility of knowing while it is almost certain on one’s evidence that one does not know, we may misclassify those cases as counterexamples to principles that are in fact sound.

1. It is useful to explore the issues within a framework taken from the standard possible worlds semantics for epistemic logic, introduced by Hintikka (1962). Such a formal framework keeps us honest, by making it straightforward to check whether our descriptions of examples are consistent and what their consequences are, and by facilitating the identification of structurally appropriate models. Of course, we must also consider whether the mathematical models we use are realistic on their intended epistemic interpretation in the relevant respects. It will be argued below that the respects in which they are idealized are consistent with the uses to which the models are here being put. They resemble physicists’ idealization of planets as point masses for purposes of some calculations.

We recall some basic features of possible worlds models for epistemic logic. For present purposes, we can make two convenient simplifications. First, we need only consider one agent at a time. Although important analogues of the failure of the KK principle arise for the interpersonal case too (Williamson 2000: 131-4), we can ignore such complications here. Second, we can ignore the strictly semantic aspect of possible world models, by discussing propositions rather than sentences; the resultant structures are *frames*.

Given those simplifications, a frame is just an ordered pair <*W*, *R*>, where *W* is a set and *R* a binary relation on *W* (a set of ordered pairs of members of *W*). Informally, we think of *W* as the set of relevant worlds or maximally specific states of affairs. Correspondingly, we think of the subsets of *W* as propositions; a proposition $p \subseteq W$ is true in a world *w* if and only if $w \in p$. Obviously the conjunction of two propositions is their set-theoretic intersection, the negation of a proposition is its complement in *W*, and so on. It does not matter for these purposes whether the worlds could really have obtained; their significance may be epistemic rather than metaphysical.

Whereas the account of propositions is encoded in the frame by *W*, all the epistemology is packed into *R*. Informally, we think of *R* as a relation of epistemic possibility between worlds: a world *x* is epistemically possible in a world *w* (*wRx*) if and only if, for all one knows in *w*, one is in *x*, in other words, whatever one knows in *w* is true in *x* (where 'one' refers to the relevant agent and the present tense to the relevant time). For $p \subseteq W$, we define:

$$Kp = \{w \in W\colon \ \forall \ x \in W\ (wRx \rightarrow x \in p)\}$$

Informally, *Kp* is to be the proposition that one knows *p*. Thus one knows *p* if and only if *p* holds in every state of affairs consistent with what one knows. As an attempt to analyse knowledge in other terms, that would be circular: but its intended purpose is more modest, simply to unpack the account of knowledge encoded in the frame by *R*.

On its intended reading, the definition of *K* presupposes that one knows something if and only if it is true in all epistemic possibilities for one, that is, in all worlds in which whatever one knows is true. Notoriously, this involves assumptions about the agent's logical omniscience. It does so in two ways. First, any treatment of *K* as a function from propositions to propositions automatically requires that if *p* is the same proposition as *q* then *Kp* is the same proposition as *Kq*. Since propositions are being treated as sets of worlds, this means that if *p* is true in the same worlds as *q*, then *Kp* is true in the same worlds as *Kp*. Moreover, since truth in a world respects the usual truth-functions — a conjunction is true in a world if and only if all its conjuncts are true in that world, the negation of a proposition is true in a world if and only if the original proposition is not true in that world, and so on — truth-functionally equivalent propositions are identical, so both are known or neither is; in particular, every truth-functional tautology is known if any is. That first form of logical omniscience is independent of the specifics of the right-hand side of the definition. The second form depends on those specifics, but unlike the first applies whenever premises $p_1, \ldots, p_n$ entail a conclusion q ($p_1 \cap \ldots \cap p_n \subseteq q$), even if the conclusion is not identical with any of the premises; it says that, in those circumstances, knowing the premises entails knowing the conclusion ($Kp_1 \cap \ldots \cap Kp_n \subseteq Kq$), for if each premise is true in all worlds in which whatever one knows is true, the conclusion is also true in all worlds in which whatever one knows is true. Neither closure principle for knowledge is conditional on the agent's having carried out the relevant deduction. The idea is rather that knowing that 2 + 2 = 4 is, *ipso facto*, knowing Fermat's Last Theorem — or, to take an example that strictly concerns truth-functional equivalence, knowing $p \rightarrow p$ is, *ipso facto*, knowing $((q \rightarrow r) \rightarrow q) \rightarrow q$.

A few philosophers, such as Robert Stalnaker (1999: 241-73), are drawn to an heroic defence of logical omniscience as a surprising literal truth — or rather as a not surprising one, since by its own lights we knew it all along. On more standard views of propositional attitudes, logical omniscience is an extreme idealization. Someone can know one tautology without knowing them all. For present purposes we may assume the idealization to be harmless, for if the total evidence of a logically omniscient agent can be radically misleading about their epistemic position, our own abject failure of logical omniscience will not save us from the same fate. Roughly speaking, the crucial point turns out to be that bad logic cannot compensate for bad eyesight. We will return to this issue towards the end of the paper.

An effect of full logical omniscience is that for each world w there is the strongest proposition known by the agent in w, $R(w)$, in the sense that it is known in w and entails every proposition that is known in w. We can define $R(w) = \{x \in W\text{: } wRx\}$. Then a proposition p is known in w if and only if p follows from $R(w)$; more formally, $w \in Kp$ if and only if $R(w) \subseteq p$, by definition of K. $R(w)$ encapsulates what one knows in w.

One constraint on the epistemic possibility relation will hold for all the models of interest below. R is reflexive (wRw); every world is epistemically possible in itself, because knowledge entails truth: whatever one knows in a world is true in that world, for all one knows in w one is in w. Consequently, $w \in R(w)$ for any world w, and $Kp \subseteq p$ for any proposition p.

Counterexamples to the KK principle have a well-known formal structure in such frames for epistemic logic. Here is a toy example. W is a three-member set $\{x, y, z\}$:

x ————————————— y ————————————— z

In the diagram, R holds between worlds just in case they are identical or neighbours; thus R is both reflexive and symmetric. R is not transitive, because xRy and yRz but not xRz. The strongest things known in each world are these: $R(x) = \{x, y\}$; $R(y) = \{x, y, z\}$; $R(z) = \{y, z\}$. In effect, if one is at one of the endpoints, what one knows is that one is not at the other endpoint; if one is at the midpoint, one knows nothing non-trivial about one’s position. Now let $p = \{x, y\}$. Then $Kp = \{x\}$: one knows p in x because p is true in all worlds epistemically possible in x; one does not know p in y because p is false in z, which is epistemically possible in y. Consequently $KKp = K\{x\} = \{\}$: one does not know Kp in x because Kp is false in y, which is epistemically possible in x. Thus the KK principle fails in x, because Kp is true and KKp false there.

As is well known, the non-transitivity of R is necessary and sufficient for a frame to contain a counterexample to the KK principle. For consider any frame $<W, R>$. Suppose that R is non-transitive. Thus for some x, y, z in W, xRy and yRz but not xRz. By definition, $KR(x)$ is true in x. $KR(x)$ is not true in y, because yRz and $R(x)$ is not true in z (since not xRz). Therefore $KKR(x)$ is not true in x, because xRy. Thus the KK principle fails in x. Conversely, suppose that there is a counter-example to the KK principle in $<W, R>$, say in $x \in W$. Thus for some $p \subseteq W$, Kp is true in x and KKp false in x. By the latter, for some $y \in W$, xRy and Kp is false in y, so for some $z \in W$, yRz and p is false in z. But not xRz, otherwise Kp is false in x, contrary to hypothesis. Thus R is non-transitive.

Of course, the existence of such non-transitive frames for epistemic logic does not by itself establish that there are counterexamples to the KK principle on its intended interpretation, for it remains to be shown that these mathematical structures represent genuinely possible epistemic situations. Before we turn to such matters, however, we must first enhance the frames with probabilistic structure, so that we can model issues about the probability on one's evidence that one knows something.

2. In adding probabilities to a frame $<W, R>$, the account of evidential probability in Williamson (2000: 209-37) will be followed. We start with a prior distribution $\mathrm{Prob}_{\mathrm{prior}}$ over propositions. Thus we can take a probabilistic epistemic frame to be an ordered triple $<W, R, \mathrm{Prob}_{\mathrm{prior}}>$, where W and R are as before and $\mathrm{Prob}_{\mathrm{prior}}$ is a probability distribution defined over subsets of W.

In the frames considered in detail below, $\mathrm{Prob}_{\mathrm{prior}}$ always takes the particularly simple form of a uniform distribution over the subsets of a finite set W, in the sense that every world has the same probabilistic weight as every other world. Thus, where $|p|$ is the cardinality of $p \subseteq W$, $\mathrm{Prob}_{\mathrm{prior}}(p) = |p|/|W|$. It is not suggested that non-uniform or infinite probability distributions are in any way illegitimate. However, if a non-uniform distribution were used to illustrate the epistemic phenomena in question, they might look like artefacts of gerrymandering. Similarly, if W were infinite, the phenomena might look like paradoxes of infinity, given the complications of probability distributions over infinite sets. It is therefore best to use a uniform prior distribution over a finite space where possible, to keep the argument as above-board and straightforward as we can.

For such uniform prior distributions, every nonempty subset of W has nonzero probability. We can therefore unproblematically define prior conditional probabilities by ratios in the usual way: $\text{Prob}_{\text{prior}}(p \mid q) = \text{Prob}_{\text{prior}}(p \cap q)/\text{Prob}_{\text{prior}}(q)$ whenever q is nonempty and so $\text{Prob}_{\text{prior}}(q) > 0$.

The evidential probability of a proposition in a world w is identified with its probability conditional on one's total evidence in w. One's total evidence in w can in turn be identified with the total content of what one knows in w (Williamson 2000: 184-208). In a frame $<W, R>$, the total content of what one knows in w is just $R(w)$. Since $w \in R(w)$, $R(w)$ is always nonempty, so probabilities conditional on $R(w)$ are always well-defined in the frames of most interest. So if $\text{Prob}_w(p)$ is the evidential probability in w of a proposition p:

$$\text{Prob}_w(p) = \text{Prob}_{\text{prior}}(p \mid R(w)) = \text{Prob}_{\text{prior}}(p \cap R(w))/\text{Prob}_{\text{prior}}(R(w))$$

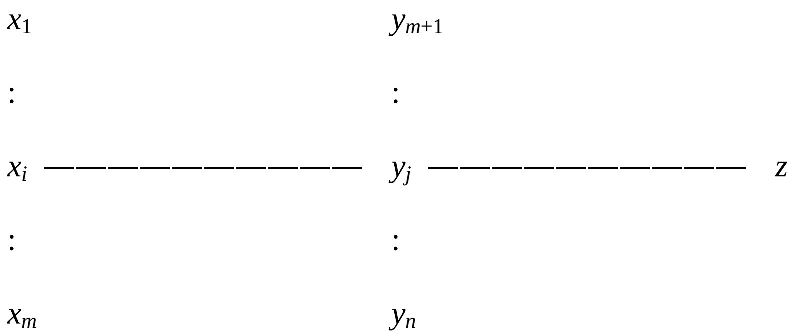
Thus in finite uniform frames, the evidential probability in w of p is simply the proportion of epistemically possible worlds in w in which p is true.

We can locate propositions about evidential probabilities in the frame. For instance, the proposition $[Pr(p) = c]$ that the evidential probability of p is the real number c is simply $\{w \in W: \text{Prob}_w(p) = c\}$, and similarly for inequalities involving evidential probabilities. Thus propositions about evidential probabilities will themselves have evidential probabilities.

Let $<W, R>$ be the three-world toy example from the previous section. As before, p is $\{x, y\}$, so Kp is $\{x\}$. Let $\text{Prob}_{\text{prior}}$ be the uniform distribution for W, so $\text{Prob}_{\text{prior}}(\{x\}) = \text{Prob}_{\text{prior}}(\{y\}) = \text{Prob}_{\text{prior}}(\{z\}) = 1/3$. Hence $\text{Pr}_x(Kp) = ½$, since Kp is true in just one of the two worlds that are epistemically possible in x. Thus in x, even though one knows p, the probability on one’s evidence that one knows *p* is no more than 50-50. To say that the probability on one’s evidence that one knows *p* is just ½ is already to say something much worse about the epistemic status for one of the proposition that one knows *p* than merely to say that one does not know that one knows *p*.

Let $<W, R, \text{Prob}_{\text{prior}}>$ be any probabilistic epistemic frame where *W* is finite, *R* is reflexive and $\text{Prob}_{\text{prior}}$ is uniform. Suppose that the KK principle fails in the frame. So for some $w \in W$ and $p \subseteq W$, *Kp* is true in *w* while *KKp* is false in *w*. Hence for some $x \in R(w)$, *Kp* is false in *x*. Since $\text{Prob}_{\text{prior}}(\{x\}) > 0$, $\text{Prob}_w(Kp) < 1$. Thus wherever the KK principle fails in such models, one knows something although the probability on one’s evidence that one knows it is less than 1. By contrast, if the KK principle holds in a frame, if *Kp* is true in *w* then *KKp* is true in *w*, so $R(w) \subseteq Kp$, so $\text{Prob}_w(Kp) = 1$: whenever one knows something, the probability on one’s evidence that one knows it is 1. Indeed, in such frames knowing *p* is equivalent to its having probability 1 on one’s evidence; $Kp = [Pr(p)=1]$. Thus the KK principle is equivalent in these circumstances to the principle that if the evidential probability of *p* is 1, then the evidential probability that the evidential probability of *p* is 1 is itself 1.

In frames where *W* is finite, *R* is reflexive and $\text{Prob}_{\text{prior}}$ is uniform, how low can $\text{Prob}_w(Kp)$ be when *Kp* is true in *w*? When one knows something, how low can the probability that one knows it be on one’s evidence? At least formally, the probability can be any rational number whatsoever strictly between 0 and 1, so it can be arbitrarily close to 0. To see this, let m/n be any rational number such that $0 < m/n < 1$, where *m* and *n* are positive integers, so $0 < m < n$. We construct a model of the required kind with some worlds in which *Kp* is true while the evidential probability of *Kp* is m/n. The idea is simple: take the three-world toy model used to illustrate the failure of the KK principle in section 1, multiply the world x by m, yielding m mutually indiscernible copies, and the world y by $n-m$, yielding $n-m$ mutually indiscernible copies. Thus $W = \{x_1, \ldots, x_m, y_{m+1}, \ldots, y_n, z\}$. R is reflexive and symmetric; for any i, i^*, j, j^* where $1 \leq i, i^* \leq m < j, j^* \leq n$: $Rx_ix_{i^*}$, Rx_iy_j, $Ry_jy_{j^*}$, and Ry_jz, but not Rx_iz; diagrammatically:



If $p = \{x_1, \ldots, x_m, y_{m+1}, \ldots, y_n\}$, for $1 \leq i \leq m$, $R(x_i) = p$. Thus $Kp = \{x_1, \ldots, x_m\}$. Consequently, while in x_i Kp is true, it is true in only m of the n epistemically possible worlds. Since $\text{Prob}_{\text{prior}}$ is uniform, $\text{Prob}_{xi}(Kp) = m/n$, as required. In particular, for $m = 1$, $\text{Prob}_{xi}(Kp) = 1/n$, which goes to 0 as n goes to infinity. By using non-uniform prior probability distributions or infinite sets of worlds we could construct similar models in which Kp is true while actually having evidential probability 0, but such refinements are unnecessary here.[1]

In the model just illustrated, m/n is the highest evidential probability that Kp attains anywhere, since $\text{Prob}_{yj}(Kp) = m/(n+1)$ and $\text{Prob}_z(Kp) = 0$. Thus the proposition *[Pr(Kp)≤m/n]* is true at every world in the model. Consequently, so is the proposition K^k*[Pr(Kp)≤m/n]*, where k is a natural number and K^k means k iterations of K (thus K^0q is *q* and $K^{k+1}q$ is K^kKq). In other words, knowing *p* is compatible not just with the probability on one's evidence that one knows *p* being close to 0, but even with knowing, and knowing that one knows, and knowing that one knows that one knows, … that the probability on one's evidence that one knows *p* is close to 0.

We have added evidential probabilities to epistemic models in a way that embodies several strong assumptions. In particular, one's total evidence was equated with the total content of one's knowledge, and probabilities on that evidence were calculated by conditionalizing a prior probability distribution on it. These assumptions are defensible (Williamson 2000), but can of course be challenged. However, someone who denied that they always hold would not be thereby committed to rejecting their present applications. For they are being used to argue that a specific phenomenon *can* occur, not that it *always* occurs. The former requires only that the relevant models *can* be instantiated by genuine epistemic phenomena, not that all genuine epistemic phenomena are similar in structure to those models. Indeed, the assumptions at issue should make it *harder*, not easier, to construct models with the target phenomenon, which involves a sort of tension between knowing and the evidential probability of knowing. For what is most distinctive about the present approach is the intimate connection it postulates between evidential probabilities and knowledge. Thus the assumptions cramp attempts to arrange the tension between them, by keeping them tightly related. By contrast, an approach that allowed more independence between evidential probabilities and knowledge would have correspondingly more scope to arrange the tension as an artefact, by varying the evidential dimension independently of the knowledge dimension or *vice versa*. Similarly, allowing non-uniform prior probability distributions or infinite sets of worlds would give far more scope for arranging odd probabilistic phenomena, for example by giving special weight to ‘bad’ worlds. If the target phenomenon occurs even under the unhelpful restrictive conditions postulated by the present approach to evidential probabilities, it is robust. A less restrictive approach could hardly rule out the models already constructed. The challenge to opponents is to motivate an approach that is *more* restrictive in some relevant way.

So far, however, we have been working at the level of formal models, without any positive argument that they represent genuinely possible epistemic situations. We now turn to that task, and provide a much more realistic, only slightly schematized description of a mundane type of epistemic situation that exemplifies the target phenomenon.

3. Imagine an irritatingly austere modernist clock. As so often happens, the designer sacrificed genuine functional efficiency for the sake of an air of functional efficiency. The clock consists of a plain unmarked circular dial with a single pointer, the hour hand, which can point at any one of *n* equally spaced, unmarked positions on the perimeter of the face. You wake up after a long sleep with no clue to the time other than the clock. You judge the time by looking at the clock, which is some distance away. Alternatively, in order to finesse the complications resulting from the movement of the hand, we could suppose that you have only a photograph of the clock, and must judge the time when it was taken.

To go into more detail, we measure distances between positions by the minimum number of steps needed to go from one to the other (clockwise or anti-clockwise). Number the positions 0, .., $n-1$ clockwise from the top. Since *n* is very large, the time period corresponding to each position is tiny, perhaps a few microseconds. For simplicity, we individuate ‘worlds’ (mutually exclusive and jointly exhaustive relevant circumstances) just by the position of the hand. Thus there are exactly n worlds w_0, .., w_{n-1}, where in w_i the hand points at position i. We measure distances between worlds by the corresponding distances between positions. An *interval* of positions corresponds to a time interval; it is a nonempty proper subset of the set of positions such that the hand goes through every position in the set without going through any position not in the set (intervals are ‘connected’). An *endpoint* of an interval is a member next to a non-member. Any interval with at least two members has two endpoints. An interval with an odd number of members has a unique *midpoint*, equidistant from its endpoints. Given the natural one-one correspondence between worlds and positions, the terms ‘interval’, ‘endpoint’ and ‘midpoint’ can be applied just as well to sets of worlds as to sets of positions.

Now imagine that you are looking at the dial from a fixed point of view equidistant from all points on the perimeter. You can make some discriminations between positions, and correspondingly between worlds, but the difference between neighbouring positions is well below your threshold of discrimination. We may assume that your capacity to discriminate between positions depends only on their relative distance; thus if world w is at least as close to world x as world y is to world z, then you can discriminate w from x only if you can also discriminate y from z (psychologically, this is doubtless an over-simplification — for example, we may be better at discriminating close to the vertical and to the horizontal than in between — but such complications are not essential to the nature of knowledge). Consequently, if you are in fact in world w, the worlds that for all you know you are in (the epistemically possible worlds) are those at most *h* steps from *w*, for some natural number *h*; *h* is greater than 0, otherwise your discrimination would be perfect. We can regard *h* as the width of the margin for error you require in order to know something in the model (Williamson 2000: 125-34); it is a constant with respect to the given model. More formally, let *R* be the epistemic accessibility relation; then for all worlds *w* and *x*, *wRx* if and only if the distance between *w* and *x* is at most *h*. Thus *R* has both reflective and rotational symmetry. For any world *w*, *R(w)* (the set of worlds epistemically possible in *w*) is not the whole of *w*, otherwise you could not make any discriminations at all. Thus *R(w)* is an interval with *w* as its midpoint and 2*h* + 1 members. For instance, if the distance *h* around the circumference is five minutes by the clock, *R(w)* corresponds to the period starting five minutes before the time *w* and ending five minutes after *w*.[2] For vividness, we will discuss your knowledge of what the time is, but under the assumption that the only source of doubt concerns the position of the hand; we assume for simplicity that the accuracy of the clock is given.

The formal epistemic model deals only in coarse-grained propositions, sets of worlds, regardless of how they are linguistically or mentally expressed. However, the model is hard to apply unless we get more specific about that. We are interested in knowledge expressible by sentences like ‘The time is 4 o’clock’, not by sentences like ‘The time is now’, even if the two sentences express the same proposition when uttered at 4 o’clock. A less direct form of cheating consists in using an indexical to exploit, to a greater or lesser degree, the position of the hand (‘The time is that’). We therefore restrict the relevant modes of temporal reference in what follows ‘The time is’ to ‘objective’ ones that do not exploit the temporal location of the thought or utterance or the spatial location of the hand. For example, *R(w)* might be expressed by a sentence like 'The time is between 3.55 and 4.05' (never mind whether it is a.m. or p.m., or what day it is).[3]

We can prove that *R(w)* is known *only* at *w*. For suppose that *R(w)* is known at a world *x*. Since *R(x)* is the strongest proposition known at *x*, $R(x) \subseteq R(w)$. But *R(x)* and *R(w)* have the same finite number of members, $2h + 1$. Thus $R(x) = R(w)$. So the midpoint of *R(x)* is the midpoint of *R(w)*; that is, $x = w$. Hence *KR(w)* is true in *w* and in no other world.

Now add evidential probabilities to the model as above, with a uniform prior distribution, $\text{Prob}_{\text{prior}}$. Since *KR(w)* is true in just one of the $2h + 1$ worlds in *R(w)*, its evidential probability in *w*, $\text{Prob}_w(KR(w))$, is $1/(2h+1)$. By increasing the number of positions round the dial while keeping your discriminatory capacities fixed, we can increase *h* without limit, and thereby make the evidential probability in *w* that one knows *R(w)* as small as desired, even though *R(w)* is in fact known in *w*.

As in section 2, the evidential probability in the model that the proposition is known is not only small, but known to be small, and known to be known to be small, and known to be known to be known to be small, and …. For since *KR(w)* is true in only one world, and for any world *x R(x)* has $2h + 1$ members, Prob*x*(*KR(w)*) is always at most $1/(2h + 1)$. Thus the proposition $[Pr(KR(w))\leq 1/(2h + 1)]$ is true in every world in the model. Consequently, the proposition $K^k[Pr(KR(w))\leq 1/(2h + 1)]$ is also true in every world. In other words, one can have any number of iterations of knowledge that the probability of *R(w)* is at most $1/(2h + 1)$.

One cannot avoid these structural results by tightening the conditions for knowledge, short of complete scepticism. For reducing the range of truths known amounts in this setting to increasing the margin for error *h*. But, given the symmetry of the situation, the argument holds for any positive margin for error — unless *h* is made so large that *R(w)* is the whole of *W*, which is in effect to say that one learns nothing by looking at the dial.[4]

Even denying the equation of evidence with knowledge would make very little difference to the argument. It would presumably involve postulating one margin for error *h* for knowledge and a distinct margin for error *h** for evidence: the worlds compatible with the total content of one's evidence in *w* would be those within a distance *h** of *w*; *h** is nonzero too, for more than one position is compatible with one's evidence. That would not affect the argument that *KR(w)* is true in no world except *w*. Hence the probability on one's evidence in *w* of *KR(w)* would be 1/(2*h** + 1). By increasing the number of positions for the dial, one can make *h** arbitrarily high, and therefore the probability on one's evidence in *w* that one knows *R(w)* arbitrarily low.

We can even generalize the argument from knowledge to rational belief (or justified belief), while keeping an independent standard for evidence (as in the previous paragraph). Unlike knowledge, rational belief is supposed not to require truth. Standardly, epistemic logic for rational belief differs from epistemic logic for knowledge just in replacing the principle that what is known is true by the principle that what it is rational to believe is consistent. For an operator for rational belief (rather than knowledge), a world *x* is accessible from a world *w* if and only if whatever it is rational for one to believe (rather than whatever one knows) at *w* is true at *x*. Technically, the constraint that the accessibility relation be reflexive is relaxed to the constraint that it be *serial*, in the sense that no world has it to no world. The effect is that the T axiom

$Kp \rightarrow p$ is weakened to the D schema $Jp \rightarrow \neg J\neg p$ (writing *J* for 'it is rational for one to believe that' or 'one is justified in believing that'). Of course, dropping the constraint that accessibility be reflexive does not imply adding the constraint that it be non-reflexive. In the present case, since we may build into the example that it is quite clear to one from general background information that one is not suffering from any illusion or systematic distortion of perception; one's only problem is the limit on one's powers of perceptual discrimination. Thus, as before, for any worlds *w, x, y* and *z* in the model, if *w* is at least as close to *x* as *y* is to *z* (in number of steps around the circumference) then *x* is accessible from *w* if *z* is accessible from *y*. As before, that implies that for some natural number *h*** (perhaps distinct from *h* and *h**), constant across the model, one world is accessible from another if and only they are at most *h*** steps apart. In particular, every world is accessible from itself, not by force of any general constraint about rational belief, but simply as a feature of this specific epistemic situation. Rational belief sometimes behaves like knowledge. Thus the structure is just as in the previous paragraph, with the upshot that it can be rational for one to believe a proposition even though it is almost certain on one's evidence that it is not rational for one believe that proposition.

If the condition for rational belief were relaxed from 'in all accessible worlds' to 'in most accessible worlds', the inference from *Jp* & *Jq* to *J*(*p* & *q*) would fail. For 'in most accessible worlds' will be equivalent to 'in at least *k* accessible worlds' for some given natural number *k* greater than *h*** and less than 2*h*** + 1. Let *w* be a world, and *p* be a subset with exactly *k* members of the set *q* of worlds accessible from *w*, including the two endpoints *x* and *y* of *q*. Thus *Jp* is true at *w*. We may assume that the total number of worlds is well over 4*h***, since otherwise one's eyesight in the model is so bad that when the hand is pointing at twelve o'clock, it is not even rational for one to believe that it is pointing between eleven o'clock and one o'clock. Then from any world other than *w*, *x* and *y* are not both accessible, so not all members of *p* are accessible, so fewer than *k* members of *p* are accessible, so it is false that most members of *p* are accessible. Thus *w* is the *only* world at which *Jp* is true, the only world at which it is rational for one to believe *p*. From here the argument proceeds as before.

The foregoing results should still hold on reasonable variations in the prior probability distribution *R(w)* that make it slightly non-uniform, for *KR(w)* will still be true only in *w*, and so its probability (the probability that one is in *w*) will still be low in *w* and *a fortiori* everywhere else too. Similarly, making the allowable space of positions for the hand continuous rather than discrete should not make much difference. One would also expect the target phenomenon often to arise in comparable ways when the epistemic accessibility relation *R* takes different forms, for example by being linear or multi-dimensional. Nor do psychologically more realistic descriptions of knowledge seem to raise the probability on one's evidence that one knows the strongest relevant proposition one can know, when one does in fact know it. Thus the target epistemic phenomenon seems robust.

Reflection suggests a generalization of the example. One key structural feature of the model is this:

(*) For all worlds w, x: $R(x) \subseteq R(w)$ only if $x = w$.

That is, shifting from one world to another (as from *w* to *x*) always opens up new epistemic possibilities as well as perhaps closing down old ones. Some worlds are close enough to *x* to be epistemically possible in *x* but not close enough to *w* to be epistemically possible in *w*. This is a plausible feature of real-life examples of inexact knowledge. As we move through logical space, our epistemic horizon moves with us. New epistemic possibilities enter our epistemic horizon as others depart. In fact, this more limited feature may do in place of (*):

(**) Some world *w* is such that for all worlds *x*: $R(x) \subseteq R(w)$ only if $x = w$.

For, given (**), *w* is still the only world in which $R(w)$ is known, so the evidential probability of $KR(w)$ will tend to be small in *w*, given the subject's imperfect powers of discrimination, even though $KR(w)$ is always true in *w*. In particular, the circular geometry of the example is not necessary for the main conclusions: it is just that the symmetry of the circle considerably simplifies the model and the arguments.

Note that if at least one verifying world *w* for (**) has *R* to a world other than itself (in other words, one is not omniscient in that world), then *R* is non-transitive. For suppose that wRx, $x \neq w$ and $R(x) \subseteq R(w)$ only if $x = w$. Then for some world *y*, xRy but not wRy, so transitivity fails.

The existence of natural structural generalizations such as (*) and (**) provides some further confirmation of the robustness of the phenomenon of knowing that is highly improbable on the subject's own evidence.[5]

4. One restrictive feature of the model in section 3 is that the width of the margin for error required for knowledge is in effect treated as beyond doubt, since it is built into the structure of the model. More specifically, since the model has only one world in which the clock hand has a given position, worlds can differ over what positions are epistemically possible for the hand only by differing over which position it in fact has. Yet it is overwhelmingly plausible that there is inexactness in our knowledge of the width of the margin for error in addition to the inexactness in our knowledge of the position of the hand. If so, then in more realistic models the worlds epistemically possible in a given world *w* will include some in which the margin for error differs slightly from that in *w*, while the position of the hand is the same. In particular, in *w* a world *x* is epistemically possible in which the margin for error is slightly less than in *w*. In such cases we may have $R(x) \subseteq R(w)$ even though $x \neq w$. Pictorially: a sphere may contain a sphere of slightly smaller radius whose centre is a slight distance from the centre of the first sphere. Then whatever is known in *w* is also known in *x*. In such cases, (*) and even (**) may fail.[6]

To construct models with a variable margin for error is not hard. But doing so without making *ad hoc* choices is much harder. In effect, one must specify higher-order margins for error distinct from the first-order margins for error. There is no obvious non-arbitrary way of determining the relation between the widths of the margins at different orders. By contrast with the simpler case in section 3, it is not clear which models one should be considering. As a consequence, it is harder to distinguish mere artefacts of the model from more significant results.

Nevertheless, in a setting with variable margins for error, one can still give an informal argument for a conclusion similar to that already reached in the case of constant margins. Let *H(w)* be the strongest proposition known in *w* about the position of the hand (or whatever other non-epistemic fact is relevant). Thus *H(w)* may be true at worlds other than *w*; its truth-value remains constant across worlds where the position of the hand is the same, even if the epistemic facts differ. Let *h* be the first-order margin for error (the one relevant to knowledge of the position of the hand) in *w*. Thus *H(w)* is true in exactly those worlds where the distance of the hand position from that in *w* is at most *h*. Let $ME_{<w}$ be true in just those worlds in which the first-order margin of error is less than *h*, $ME_{>w}$ be true in just those worlds in which the first-order margin for error is greater than *h*, and $ME_{=w}$ be true in just those worlds in which the first-order margin for error is equal to *h*. These three possibilities are mutually exclusive and jointly exhaustive. Therefore, by definition of conditional probability:

$$(1)\qquad \mathrm{Prob}_w(KH(w)) = \mathrm{Prob}_w(KH(w) \mid ME_{<w}).\mathrm{Prob}_w(ME_{<w}) + \mathrm{Prob}_w(KH(w) \mid ME_{=w}).\mathrm{Prob}_w(ME_{=w}) + \mathrm{Prob}_w(KH(w) \mid ME_{>w}).\mathrm{Prob}_w(ME_{>w})$$

In any world *x* in $ME_{>w}$ some world is epistemically possible in which *H(w)* is false, because the first-order margin for error in *x* is some $k > h$, and a sphere of radius *k* cannot be contained in a sphere of radius *h*. Thus $ME_{>w}$ is incompatible with *KH(w)*, so $\mathrm{Prob}_w(KH(w) \mid ME_{>w}) = 0$. Consequently, (1) simplifies to:

$$(2)\qquad \mathrm{Prob}_w(KH(w)) = \mathrm{Prob}_w(KH(w) \mid ME_{<w}).\mathrm{Prob}_w(ME_{<w}) + \mathrm{Prob}_w(KH(w) \mid ME_{=w}).\mathrm{Prob}_w(ME_{=w})$$

Since $\text{Prob}_w(KH(w) \mid ME_{<w}) \leq 1$, (2) yields:

(3) $\text{Prob}_w(KH(w)) \leq \text{Prob}_w(ME_{<w}) + \text{Prob}_w(KH(w) \mid ME_{=w}).\text{Prob}_w(ME_{=w})$

For simplicity, we may reasonably assume that $\text{Prob}_w(ME_{<w}) = \text{Prob}_w(ME_{>w})$, that is, that the first-order margin for error is equally likely to be less or greater than its actual value. Since $\text{Prob}_w(ME_{<w}) + \text{Prob}_w(ME_{=w}) + \text{Prob}_w(ME_{>w}) = 1$, $\text{Prob}_w(ME_{<w}) = (1-\text{Prob}_w(ME_{=w}))/2$. Therefore, by (3):

(4) $\text{Prob}_w(KH(w)) \leq (1-\text{Prob}_w(ME_{=w}))/2 + \text{Prob}_w(KH(w) \mid ME_{=w}).\text{Prob}_w(ME_{=w})$

But $\text{Prob}_w(KH(w) \mid ME_{=w})$ is in effect the probability of *KH(w)* in the case considered previously of a constant margin for error (*h*). From that case we have at the very least that $\text{Prob}_w(KH(w) \mid ME_{=w}) < ½$. Consequently, by (4):

(5) $\text{Prob}_w(KH(w)) \leq (1-\text{Prob}_w(ME_{=w}))/2 + \text{Prob}_w(ME_{=w})/2 = ½$

In other words, although you in fact know *H(w)* in *w*, it is no more probable than not on your evidence in *w* that you know *H(w)*.

Even if we slightly relax the simplifying assumption that $\text{Prob}_w(ME_{<w}) = \text{Prob}_w(ME_{>w})$, the probability on the evidence in *w* that *H(w)* is known will not rise significantly above evens. Indeed, the probability may well be close to zero. For if the width of the first-order margin for error varies only slightly (as a proportion of *h*) over the worlds epistemically possible in *w*, then $\text{Prob}_w(KH(w) \mid ME_{<w})$ will be close to $\text{Prob}_w(KH(w) \mid ME_{=w})$. Therefore, by (2), $\text{Prob}_w(KH(w))$ will be at most only slightly greater than $\text{Prob}_w(KH(w) \mid ME_{=w}).\text{Prob}_w(ME_{<w}) + \text{Prob}_w(KH(w) \mid ME_{=w}).\text{Prob}_w(ME_{=w}) = \text{Prob}_w(KH(w) \mid ME_{=w}).\text{Prob}_w(ME_{\leq w})$. But, as noted above, $\text{Prob}_w(KH(w) \mid ME_{=w})$ is in effect the probability of *KH(w)* in the case already considered of a constant margin for error. That probability goes to zero as the number of hand positions increases. Hence $\text{Prob}_w(KH(w))$ may well be close to 0 even when the width of the margin for error varies. But even without that stronger conclusion, the result of the informal argument is enough for present purposes. Uncertainty about the width of the margin for error does not undermine the possibility of knowing something without its being probable on one's evidence that one knows it.

5. We can elaborate the clock example with variable margins for error by filling in more of the internal mechanism in a way that reinforces the original moral. Suppose that the hand has, in addition to its real position, an apparent position constituted by a state of the agent. We can model each world as an ordered pair <*j*, *k*>, where *j* is the real position of the hand and *k* is its apparent position, *j* and *k* both being drawn from the same set of positions. Correspondingly, we will speak of real and apparent times in the obvious way (recall that the accuracy of the clock is given). This is of course still an over-simplification, but that is the nature of modelling. We do not assume that at <*j*, *k*> you *believe* that the time is *k*. You may just believe that it is within a particular time interval that contains *k*. We assume that that the same general perceptual mechanism is operating

at all worlds, and that the real position of the hand is not the only factor in determining its apparent position, indeed, the perceptual mechanism puts no upper limit on the size of the ‘error’, the distance between j and k. All pairs are treated as worlds. However, a pair $<j, k>$ where j and k are very far apart may still be suitably remote from the actual world, in ways to be explained.

To define the model, we must specify when $<j, k>R<j^*, k^*>$. To keep things simple, and to make a generously unrealistic concession to an old-fashioned foundationalist conception of knowledge, we may even treat the agent as omniscient about the appearances they are currently entertaining, by the stipulation that $<j, k>R<j^*, k^*>$ only if $k = k^*$, in other words, if the apparent time is k then the agent knows that the apparent time is k. Thus all that remains is to specify when $<j, k>R<j^*, k>$.

If $j = k$, the appearance is perfectly accurate, but of course it does not follow that the agent knows exactly what time it is. As already noted, the agent may not even have a belief as to exactly what time it is. However, on general grounds of symmetry we may assume that at $<j, j>$ what the agent knows about the time is that it is within a given number h of steps from j. The value of h will be determined by features of the visual mechanism, such as the long-run probability of a given distance between the real and apparent positions of the hand, which for simplicity we assume to be invariant under symmetries (rotations and reflections) of the circle. Let $d(j, j^*)$ be the distance between positions j and j^* as measured by the number of steps on the shortest path round the circumference between them (so d is a metric in the standard topological sense). Then $<j, j>R<j^*, j>$ if and only if $d(j, j^*) \leq h$. Since the agent does not know at $<j, j>$ exactly what time it is, $h \geq 1$.

Now what remains is to specify when $<j, k>R<j^*, k>$ if $j \neq k$. The natural assumption is that what the subject knows at $<j, k>$ about the time is in effect that it is within a given distance of k, the known apparent time. When $j = k$ the distance is h. As the distance between j and k increases, the required maximum distance of j^* from k will have to increase at least as fast, at least eventually, otherwise $<j^*, k>$ would sometimes be inaccessible from $<j, k>$ when $j^* = j$, so that R would be non-reflexive, contradicting the factiveness of knowledge. The obvious way to combine this desideratum with the condition already fixed for the case when $j = k$ is by setting the required maximum distance of j^* from k as $h + d(j, k)$. The greater the distance between real and apparent positions, the less the agent knows.

We have thus arrived at the rule that $<j, k>R<j^*, k^*>$ if and only if $k = k^*$ and $d(j^*, k) \leq h + d(j, k)$. By contrast, when one defines an accessibility relation R_B for (blameless) belief rather than knowledge in such a model, one will presumably make it independent of the real position j. For example, $<j, k>R_B<j^*, k^*>$ if and only if $k = k^*$ and $d(j^*, k) \leq h$: what one believes in $<j, k>$ is true in just those worlds in which the real position is close enough to its apparent position in $<j, k>$ and its apparent position is exactly the same as in $<j, k>$. Given those definitions of R and R_B, knowledge entails belief, for $R_B(<j, k>) \subseteq R(<j, k>)$, so at $<j, k>$ if the agent knows p then $R(<j, k>) \subseteq p$, so $R_B(<j, k>) \subseteq p$, so the agent believes p. In this simplified model, whenever the apparent position matches the real position, belief coincides with knowledge: $R_B(<j, j>) = R(<j, j>)$. Any failure of match between appearance and reality generates some beliefs that fail to constitute knowledge, for when $j \neq k$, $d(j, k) > 0$ so not $R(<j, k>) \subseteq R_B(<j, k>)$, so $R_B(<j, k>)$ is believed but not known at $<j, k>$. When the failure of match is not too great, it even generates true beliefs that fail to constitute knowledge, for $<j, k> \in R_B(<j, k>)$ when $0 < d(j, k) \leq h$, so $R_B(<j, k>)$ is true at $<j, k>$: all the agent's beliefs there are true, but not all of them constitute knowledge.

The extra error term $d(j, k)$ in the definition of the accessibility relation for knowledge but not for belief is exactly what makes the former but not the latter factive in this model. R is reflexive because $d(j, k) \leq h + d(j, k)$. R_B is non-reflexive because $d(j, k) > h$ when the mismatch between real and apparent positions is too great; in that case the agent falsely believes $R_B(<j, k>)$ at $<j, k>$. This difference is one manifestation of the way in which what is known at $<j, k>$ ($R(<j, k>$) depends on the real position j while what is believed ($R_B(<j, k>)$) does not. But it is not the only manifestation. For, as just seen, it also generates cases of true belief that fail to constitute knowledge. Moreover, the beliefs are justified, at least in the sense of being blameless, since what is believed is exactly what would be known if the reality matched the given appearance (($R_B(<j, k>) = R(<k, k>)$). Thus they are Gettier cases: justified true beliefs that are not knowledge (in the relevant sense of 'justified'). Where the failure of match is not too great, there are no associated false beliefs, so the cases are more similar to fake barn cases than to the original Gettier cases (Gettier 1963, Goldman 1976). That the model correctly predicts such distinctive features of knowledge is some confirmation that it is on the right lines. The extra error term $d(j, k)$ makes knowledge depend on reality for more than just truth.

As in the previous model, the accessibility relation for knowledge is non-transitive, so the KK principle fails. For instance, if $d(j, j^*) = h$, $d(j^*, j^{**}) = 1$ and $d(j, j^{**}) = h + 1$, then $<j, j>R<j^*, j>$ since $d(j, j^*) = h = h + d(j, j)$, and $<j^*, j>R<j^{**}, j>$

since $d(j, j^{**}) = h + 1 \leq 2h = h + d(j^*, j)$, but not $\langle j, j\rangle R\langle j^{**}, j\rangle$ since $d(j^{**}, j) = h + 1 > h = h + d(j, j)$.

What is new is that the accessibility relation for knowledge is also non-symmetric, unlike those in the previous models. For instance, if *j* is more than *h* steps from *k*, then $\langle j, k\rangle R\langle k, k\rangle$ (because $d(k, k) = 0 \leq h + d(j, k)$) but not $\langle k, k\rangle R\langle j, k\rangle$ (because $d(k, j) > h = h + d(k, k)$). We can interpret the example thus. The world $\langle k, k\rangle$ is a good case, in which appearance and reality match. The world $\langle j, k\rangle$ is a corresponding bad case, in which the appearance is the same as in the good case but drastically fails to match reality. The good case is accessible from the bad case: if you are in the bad case, everything you know is true in the good case. The bad case is inaccessible from the good case: if you are in the good case, something you know is false in the bad case. Sceptical scenarios involve just such failures of symmetry. Everything you know in the sceptical scenario is true in ordinary life, but something you know in ordinary life is false in the sceptical scenario (see Williamson 2000: 167, 226). The non-symmetry affects what logical principles hold in the model. For example, symmetry corresponds to the B principle that if something obtains then you know that for all you know it obtains ($p \rightarrow K\neg K\neg p$). That principle holds in the previous models but fails in this one. For example, if you are in the bad case, it does not follow that you know that for all you know you are in the bad case. Rather, if you are in the bad case, then for all you know you are in the good case, in which you know that you are not in the bad case.

Just as before, cases occur of very improbable knowing. For instance, $R(\langle j, j\rangle)$ comprises just the $2h + 1$ worlds of the form $\langle j^*, j\rangle$ such that $d(j^*, j) \leq h$. But whenever $j \neq j^*$, $R(\langle j^*, j\rangle)$ comprises more than $2h + 1$ worlds of the form $\langle j^{**}, j\rangle$ such that

$d(j^{**}, j) \leq h + d(j^*, j) > h$; thus not $R(<j^*, j>) \subseteq R(<j, j>)$, so $KR(<j, j>)$ is false at $<j^*, j>$. Consequently, although $KR(<j, j>)$ is true at $<j, j>$, $\text{Prob}_{<j, j>}(KR(<j, j>)) = 1/(2h + 1)$. In the world $<j, j>$, one knows the proposition $R(<j, j>)$, but it is almost certain on one's evidence that one does not know that proposition.

The evidential improbability of knowing $R(<j, j>)$ is reflected in one's failure to believe that one knows it. For if $1 \leq d(j^*, j) \leq h$ then $<j^*, j>$ is doxastically accessible from $<j, j>$, although in $<j^*, j>$ one does not know $R(<j, j>)$, so in $<j, j>$ one does not believe that one knows $R(<j, j>)$. Of course, in $<j, j>$ one also lacks the false belief that one does not know $R(<j, j>)$, for since $<j, j>$ is doxastically accessible from itself all one's beliefs in $<j, j>$ are true. Since in $<j, j>$ one knows $R(<j, j>)$, and knowledge entails belief, one believes $R(<j, j>)$. Thus one violates the contentious axiom that whenever one believes p, one believes that one knows p.[7]

In this model, one's beliefs depend only on what one is omniscient about, the appearances (for $R_B(<j, k>) = \{<j^*, k>: d(j^*, k) \leq h$, which is independent of j), so one is omniscient about them too. Thus if one believes p, one knows that one believes p, and if one fails to believe p, one knows that one fails to believe p. In particular, therefore, in $<j, j>$ one knows that one both believes $R(<j, j>)$ and fails to believe that one knows $R(<j, j>)$. A fortiori, one knows that one fails to know that one knows $R(<j, j>)$. That is self-conscious modesty. It is not irrationality, for, as already noted, in $<j, j>$ one believes only those things that one knows.

What is the relation between this new model and the one in section 3? We can regard the new worlds as refinements of the old ones, or conversely the old worlds as equivalence classes of the new ones. More specifically, let $|<j, k>|$ be the equivalence class of $<j, k>$, so $|<j, k>| = |<j^*, k^*>|$ if and only if $j = j^*$ (the identity of the old world depends on the real position of the pointer, not on its apparent position). We define a natural accessibility relation $R_{||}$ on the equivalence classes by the rule that $wR_{||}w^*$ if and only if each member of w has access (R) to at least one member of w^* (if you are somewhere in w, then for all you know you are somewhere in w^*). Then $R_{||}$ turns out to coincide with the old accessibility relation on the old worlds, in the sense that $<j, k>R_{||}|<j^*, k^*|$ when and only when $d(j, j^*) \leq h$. For suppose that $d(j, j^*) \leq h$. A representative member of $|<j, k>|$ is $<j, k^{**}>$. But by the triangle inequality for d (one of the axioms for a metric space):

$$d(j^*, k^{**}) \leq d(j^*, j) + d(j, k^{**}) = d(j, j^*) + d(j, k^{**}) \leq h + d(j, k^{**})$$

Hence $<j, k^{**}>R<j^*, k^{**}>$. But $<j^*, k^{**}> \in |<j^*, k^*>|$, so $<j, k>R_{||}|<j^*, k^*|$, as required. Conversely, suppose that $d(j, j^*) > h$. A representative member of $|<j^*, k^*>|$ is $<j^*, k^{**}>$., If $<j, j>R<j^*, k^{**}>$ then $d(j, j^*) \leq h + d(j, j) = h$, which is a contradiction, so not $<j, j>R<j^*, k^{**}>$. But $<j, j> \in |<j, k>|$, so not $<j, k>R_{||}|<j^*, k^*|$, as required. In this sense, we have recaptured the old model as a coarsening of the new one.

Of course, the new model does not reproduce every feature of the old one. For instance, as already seen, the B principle no longer holds. Rather, the new model corrects some unrealistic features of its predecessor. In the old model, the set of epistemically accessible worlds around a world w forms an interval of constant size $2h + 1$, no matter where w is in the model. That corresponds to the assumption that it is always known exactly *how much* is known, even though it is never known exactly *what* is known. That assumption is obviously a gross idealization. In the new model, the set of epistemically accessible worlds around a world $<j, k>$ forms an interval of size $2(h + d(j, k)) + 1$, which varies with the degree of mismatch between appearance and reality, which is at least somewhat more realistic. Unlike the old model, the new one represents the possibility of illusion and error as well as of limited discrimination. Nevertheless, with all these extra refinements, and perfectly known appearances, the new model still predicts cases of knowing while it is virtually certain on one's evidence that one does not know. The phenomenon is robust.

For simplicity, the available apparent positions of the hand have been assumed to coincide with the available real positions. We could lift this assumption and suppose that fewer apparent positions than real positions are available. This would complicate the arguments but require only minor refinements of their conclusions.

Philosophers of an internalist bent tend to restrict one's evidence to appearances. In this model, the effect is that a world $<j^*, k^*>$ is consistent with one's internalist evidence in $<j, k>$ if and only if $k^* = k$. Thus any position whatsoever for the pointer is consistent with one's internalist evidence in any world. On that view, cases of improbable knowing become even more extreme, since $<j, j>$ is the only world in the model in which $R(<j, j>)$ is known: making more worlds consistent with the evidence merely drives down the probability of $KR(<j, j>)$ on the evidence. A more radical internalist move would be to make $<j^*, k^*>$ epistemically accessible from $<j, k>$ whenever $k^* = k$. But that would yield total scepticism about the external world. For let p be any proposition purely about the external world, in the sense that whether it is true in a world $<j^*, k^*>$ depends only on the real position j^* and not on the apparent position k^*. Then p is known in a world $<j, k>$ only if it is trivial, that is, true in every world in the model. For if p is known in $<j, k>$ by the radical internalist standard then p is true in any world epistemically accessible by that standard from $<j, k>$, so in particular p is true in $<j^*, k>$; by hypothesis the truth of p in $<j^*, k>$ depends only on j^*, not on k, so p is also true in $<j^*, k^*>$. But j^* and k^* were arbitrary, so p is true in every world in the model, as claimed.

A still more realistic version of the model would reject the assumed omniscience about one's own beliefs and appearances. The inner world is not much easier to know than the outer. I find it at least as hard to introspect the time according to my phenomenal clock as to see the time according to a real clock. Such refinements do nothing to undermine the phenomenon of improbable knowing. Moreover, they reintroduce the phenomenon of improbable rational belief, as discussed in section 3.

6. Examples of the kind considered in previous sections make trouble for accounts of propositional justification on which a sufficient condition for having such justification to believe p is that the probability of p on one's evidence exceeds a threshold less than 1. For in such cases one would then have propositional justification to believe the Moore-paradoxical conjunction $p \& \neg Kp$. Consider a world w in which one knows p but $\mathrm{Prob}_w(Kp)$, the probability on one's evidence in w that one knows p, is less than $1/n$, so $\mathrm{Prob}_w(\neg Kp) > (n-1)/n$. Since one knows p, $\mathrm{Prob}_w(p) = 1$. By elementary probability theory, it follows that $\mathrm{Prob}_w(p \& \neg Kp) > (n-1)/n$. Thus by letting n go high enough, we can find a case in which the probability on one's evidence of the Moorean conjunction exceeds the given threshold and the supposedly sufficient condition for justification is met. But it is not plausible that one can have justification to believe a conjunction of the form of "It's raining and I don't know that it's raining". In effect, this point is a sort of abstract generalization of the objection to probabilistic acceptance rules from Moore

paradoxes in lottery cases (“My ticket won’t win and I don’t know that it won’t win”), but it avoids their reliance on the specific assumption that one is not in a position to know that one’s ticket won’t win.

However, the phenomenon at issue does not merely raise problems for particular philosophical views. More importantly, it raises problems for rational agents as such, not just for those rational agents who choose to theorize about rational agents. Suppose, for instance, that it is rational for Hamlet to perform some action *A* if he knows *p* and not rational for him to perform *A* otherwise (given the relevant background circumstances, such as his other attitudes), and that those facts about rationality are clear to him.[8] Suppose also that although Hamlet knows *p*, it is almost certain on his evidence that he does not know *p*. Should Hamlet do *A*? Since he knows *p*, it is rational for him to do *A*. However, since it is almost certain on his evidence that he does not know *p*, it is almost certain on his evidence that it is not rational for him to do *A*. It is therefore very tempting to say that after all it is not rational for him to do *A*. But to say that would be to contradict the conditions of the example. Thus there is pressure to say instead that the example cannot really arise, that if it is almost certain on one’s evidence that one does not know *p* then one does not really know *p*. We have already seen that such pressure should be resisted, since it will not stop short of scepticism. Nevertheless, its effect when we consider particular cases may be to make us withdraw true knowledge ascriptions, under the false impression that they have been defeated by the negative evidence.

The difficulty cannot be met by insisting that what it is rational for Hamlet to do depends on what it is rational for him to believe, not on what he knows. For we have already seen that parallel to the phenomenon of improbable knowing is a phenomenon of improbable rational believing.

By a parallel argument, just as it can be rational for one to do *A* even though it is extremely improbable on one's evidence that it is rational for one to do *A*, so it can happen that one *should* do *A*, even though it is extremely improbable on one's evidence that one should do *A*.

Of course, we have no general difficulty with the idea that a claim may be true even though it is almost certain on someone's evidence that it is false — for example, when the claim states the actual outcome of a long sequence of unobserved coin tosses. What we find harder to accept is the possibility of the same combination when the claim ascribes knowledge to the very person whose evidence is in question at that very time. Their strong evidence that they do not know *p* seems incompatible with whatever sort of reliance on *p* is mandated by their knowing *p*. But the apparent incompatibility is an illusion, which can seriously distort our assessment of particular ascriptions of knowledge and with it our epistemological theorizing.

The phenomenon that we find hard to accept is an extreme case of anti-luminosity, the failure of non-trivial states to satisfy the constraint that whenever one is in them one is in a position to know that one is in them (Williamson 2000: 93-113). In evidential terms, when luminosity fails one is in a state *S* even though it is not certain on one's evidence that one is in *S*. In the present case, *S* is the state of knowing *p*. But the phenomenon at issue is stronger than a mere failure of luminosity, since it involves being in *S* even though it is almost certain on one's evidence that one is *not* in *S*.

If we do resist the sceptical pressure, and acknowledge the possibility of the phenomenon, then we must regard Hamlet's problem above as a genuine problem for him, an instance of a more general practical problem for agents acting on inexact knowledge, not as something to be dissolved by epistemological redescription. The flat-footed solution is clear from the conditions of the example: it is rational for Hamlet to do *A* if and only if he knows *p*, and he does know *p*, so it *is* rational for him to do *A*. The trouble is that he has strongly misleading evidence about those facts. But that does not mean that they are not facts after all; it just means that he has a good excuse — not justification — for not having done *A*.[9] Hamlet is in a bad predicament, but why should we expect an epistemological theory to tell us that rational agents cannot get into bad predicaments, or trust it if it does so? Decision theory cannot show that such epistemic phenomena occur; at best it can help us act in a world where they do occur.

A related difficulty for rational agents is that in any frame in which one can know without knowing that one knows, or not know without knowing that one does not know, failures occur of the synchronic reflection principle that one's evidential probability for *p*, conditional on one's evidential probability for *p*'s being *c*, is itself *c*. Although the details are slightly more complicated, a similar situation holds for rational belief in place of knowledge, in a sense in which there are rational false beliefs (see Appendix).

We must check these general comments by examining a specific case in more detail. That is the business of the following final section.

7. Competent deduction is a way of extending our knowledge, as in mathematical reasoning. A natural formulation of the underlying principle is this form of *multi-premise closure*:

MPC If one believes a conclusion on the basis of competent deduction from premises each of which one knows, one knows the conclusion.

For present purposes we can leave the phrase 'on the basis of competent deduction' vague, and assume that something in the spirit of MPC is correct.[10]

MPC faces a strong challenge from cases of reasoning from many premises, each with a small independent risk of error, where those risks combine into a large risk of error for the conclusion. Such situations arise most simply when the conclusion is the conjunction of the premises, since then the conclusion entails the premises and any error in a premise involves an error in the conclusion; the deduction of the conclusion from the premises also takes a particularly elementary form. Suppose that each premise is true, so the conclusion is true too. We may further suppose that, according to ordinary standards, one knows each premise. For example, each premise may state a separate matter of particular historical fact that one has carefully checked. To deny that one knows a given premise would look like extreme scepticism. From those propositions one has competently deduced their conjunction and believes it on that basis. By MPC, one thereby knows the conclusion. Nevertheless, much past experience may show that there is a small non-zero rate of error for such carefully checked historical claims, on the basis of which it is almost inevitable that the long conjunction contains several false conjuncts.

Thus one’s belief in the conjunction may seem to have too strong a risk of error to constitute knowledge. This is a variant of the Preface Paradox, adapted to knowledge.

On a fallibilist conception of knowledge, it seems, the risk of error for any one premise may be within the threshold for knowledge, while the risk of error for the conclusion is outside the threshold; thus MPC fails. By contrast, on an infallibilist conception, knowledge requires that there be no risk of error, rather than at most a small risk, so MPC may hold; but such infallibilism seems to lead to scepticism. Alternatively, one might try to preserve MPC by postulating some sort of variation across contexts of utterance in the reference of ‘know’, or some loss of knowledge of the premises through the very act of deducing the conclusion.[11]

Once we recognize the phenomenon of knowing when it is almost certain on one’s evidence that one fails to know, we can see a possible diagnosis of the problem cases for MPC which allows us to keep MPC while neither falling into scepticism nor postulating truth-conditions for ‘knowledge’-ascriptions of any non-standard kind. One does indeed know each premise, without knowing that one knows it. Since one believes the conclusion on the basis of competent deduction from the premises, by MPC one also knows the conclusion, although without knowing that one knows it. For each premise, it is very probable on one’s evidence that one knows it. However, it is very improbable on one’s evidence that one knows every premise. Given that one knows the conclusion (the conjunction) only if one knows every premise, it is very improbable on one’s evidence that one knows the conclusion. Since we are tempted to conceive knowing as at least somewhat luminous, we are tempted to deny that one knows the conclusion.[12]

We can give a formal epistemic model of that description. As noted in section 1, such a model automatically validates not just MPC but logical omniscience, an unrealistically strong version of multi-premise closure that does not even require the agent to have carried out the relevant deductions, competently or otherwise. As a positive argument for MPC, it might well be accused of both begging the question and proving too much. However, its role here is different. It is being used to defuse an objection to MPC, by showing that even on the assumption of the strongest possible version of multi-premise closure, one would predict the occurrence of epistemic phenomena that it is very tempting to misinterpret along the lines of the objection as counter-examples to multi-premise closure. In particular, if we treat ascriptions of knowledge as defeated by a low probability of knowing, but not by a high probability short of 1 of knowing, on the subject's evidence, then we shall tend to judge that the subject knows each conjunct without knowing the conjunction, even though the conditions for MPC are satisfied; we are deceived by the false appearance of a counterexample to MPC.[13]

Here are the details of such a model.[14] For worlds we use n-tuples of numbers drawn from the set $\{0, 1, \ldots, 2k\}$, where n is the number of premises (conjuncts) and k is a large natural number. Thus there are $(2k+1)^n$ worlds. The n components of a world represent its locations on n independent dimensions of a state space. The ith dimension is the one relevant to the ith premise p_i. Let the ith component of the n-tuple w be w_i, and the world just like w except that its ith component is m be $w[i|m]$, so $w[i|m]_i = m$ and $w[i|m]_j = w_j$ if $i \neq j$. Let p_i be true in w if and only if $w_i > 0$. Let x be epistemically possible in w (wRx) if and only if for all i, $|w_i – x_i| \leq k$, that is, w and x do not differ by 'too much' in any of their respective components. In effect, a margin for error is applied to each of the n dimensions separately. The relation R is obviously reflexive and symmetric. We can easily check that for any world w, p_i is known (Kp_i is true) in w if and only if $w_i > k$.[15] Similarly, we can check that p_i is known to be known (KKp_i is true) in w if and only if $w_i > 2k$; hence p_i is not known to be known (KKp_i is not true) in any world in this model. In particular, in the world $<2k, \ldots, 2k>$, each premise p_i is known and none is known to be known.

As usual, the prior probability distribution is uniform: for any world w, $\text{Prob}_{\text{prior}}(\{w\}) = 1/(2k+1)^n$. We must check that the model makes the n dimensions probabilistically independent. For any given i, a proposition q is *i-based* if and only if for all worlds x and y, if $x_i = y_i$ then q is true in x if and only if q is true in y ($1 \leq i \leq n$). That is, whether an i-based proposition is true in a world depends only on the ith component of that world. In particular, p_i is an i-based proposition. Obviously, the negation of any i-based proposition is also i-based, as is any conjunction of i-based propositions. We can also prove that whenever q is an i-based proposition, so is Kq.[16] Thus Kp_i and KKp_i are also i-based propositions. Then we can prove that whenever for each i q_i is an i-based proposition, $q_1, \ldots, q_n$ are mutually probabilistically independent on the evidence in any world w, in the usual sense that the probability (on the evidence in w) of their conjunction is the product of the probabilities (on the evidence in w) of the conjuncts.[17] Although a model could have been constructed in which the evidence at some worlds establishes epistemic interdependences between the different dimensions, for present purposes we can do without such complications. In particular, $p_1, \ldots, p_n$ are mutually probabilistically independent on the evidence in any world, as are $Kp_1, \ldots, Kp_n$. But, on the evidence in the world $<2k, \ldots, 2k>$, for any given i, the probability that p_i is known is $k/(k+1)$.[18] By probabilistic independence, the probability of the conjunction $Kp_1 \cap \ldots \cap Kp_n$ is $(k/(k+1))^n$. That is the probability that each conjunct is known. But, by the logical omniscience built into the model, knowing a conjunction ($K(p_1 \cap \ldots \cap p_n)$) is equivalent to knowing each conjunct. Thus the probability on the evidence in <2*k*, …, 2*k*> that the conjunction $p_1 \cap \ldots \cap p_n$ is known is also $(k/(k+1))^n$. For fixed *k*, this probability becomes arbitrarily close to 0 as *n* becomes arbitrarily large. Thus, for suitable *k* and *n*, the world <2*k*, …, 2*k*> exemplifies just the situation informally sketched: for each conjunct one knows it without knowing that one knows it, and it is almost but not quite certain on one's evidence that one knows the conjunct; one also knows the conjunction without knowing that one knows it, and it is almost but not quite certain on one's evidence that one does *not* know the conjunction.

Of course, in some examples one's epistemic position with respect to each conjunct is better: one not only knows it but knows that one knows it. If one also knows the relevant closure principle, and knows that one satisfies the conditions for its application, one may even know that one knows the conjunction. Consequently, the probability on one's evidence that one knows the conjunction is 1. However, the previous pattern may still be repeated at a higher level of iterations of knowledge. For example, for each conjunct one knows that one knows it without knowing that one knows that one knows it, and it is almost but not quite certain on one's evidence that one knows that one knows the conjunct; one also knows that one knows the conjunction without knowing that one knows that one knows it, and it is almost but not quite certain on one's evidence that one does *not* know that one knows the conjunction. To adapt the previous model to this case, we can simply expand the set of worlds by using *n*-tuples of numbers from the set {0, 1, …, 3k} rather than {0, 1, …, 2k}, leaving the definitions of the epistemic possibility relation R and the truth-conditions of the p_i unchanged (so p_i is true in w if and only if $w_i > 0$); then <3k, …, 3k> is a world of the required type. More generally, if one uses as worlds n-tuples of numbers from the set {0, 1, …, hk}, leaving the other features of the model unchanged, then <hk, …, hk> will be a world at which one has h – 1 but not h iterations of knowledge of each conjunct, and it is almost but not quite certain on one's evidence that one has h – 1 iterations of knowledge of the conjunct; one also has h – 1 but not h iterations of knowledge of the conjunction, and it is almost but not quite certain on one's evidence that one does *not* have h – 1 iterations of knowledge of the conjunction.[19,20]

Many other variations can be played on the same theme. The general idea is this. Suppose that the epistemic status E satisfies an appropriate principle of multi-premise closure. In some situations, one attains E with respect to each conjunct, without knowing that one does so (this is possible by the anti-luminosity argument). By multi-premise closure, one also attains status E with respect to the conjunction, without knowing that one does so. Then for each conjunct it may be almost certain on one's evidence that one attains E with respect to it, even though it is almost certain on one's evidence that one does *not* attain E with respect to the conjunction. If we treat ascriptions of E as defeated by a low probability of E, but not by a high probability short of 1 of E, on the subject's evidence, then we shall tend to judge that the subject attains E with respect to each conjunct but not with respect to the conjunction, even though the conditions for multiple-premise closure principle are satisfied; we are deceived by the false appearance of a counterexample to the multi-premise closure principle.

8 The considerations of this paper raise a more general question. Knowledge claims are often thought to be defeated by various sorts of misleading evidence. In how many cases is the correct account that the subject knows, even though it is almost certain on the subject’s evidence at the time that they do not know? That is left as an open question.

Appendix: The reflection principle for evidential probability

For ease of working, we use the following notation. A *probability distribution* over a frame $<W, R>$ is a function Pr from subsets of W to nonnegative real numbers such that $Pr(W) = 1$ and whenever $X \cap Y = \{\}$, $Pr(X \cup Y) = Pr(X) + Pr(Y)$ (Pr must be total but need not satisfy countable additivity). Pr is *regular* iff whenever $Pr(X) = 0$, $X = \{\}$. Given Pr, the evidential probability of $X \subseteq W$ at $w \in W$ is the conditional probability $Pr(X \mid R(w)) = Pr(X \cap R(w))/Pr(R(w))$, where $R(w) = \{x\text{: } wRx\}$. Similarly, the evidential probability of X conditional on Y at w is $Pr(X \mid Y \cap R(w)) = Pr(X \cap Y \cap R(w))/Pr(Y \cap R(w))$. In both cases, the probabilities are treated as defined only when the denominator is positive.

The *reflection principle* holds for a probability distribution Pr over a frame $<W, R>$ iff for every $w \in W$, $X \subseteq W$ and real number c, the evidential probability of X at w conditional on the evidential probability of X being c is itself c; more precisely:

$$Pr(X \mid \{u\text{: } Pr(X \mid R(u)) = c\} \cap R(w)) = c$$

We must be careful about whether the relevant probabilities are defined. If $Pr(R(w)) = 0$ then $Pr(X \mid R(w))$ is undefined, so it is unclear what the set term $\{u\text{: } Pr(X \mid R(u)) = c\}$ means. To avoid this problem, $Pr(R(w))$ must always be positive, so that all (unconditional) evidential probabilities are defined. In particular, therefore, $R(w)$ must always be nonempty; in other words, $<W, R>$ must be *serial* in the sense that for every $w \in W$ there is an x such that wRx. For a regular probability distribution on a serial frame, all evidential probabilities are defined. Of course, it does not follow that the outer probability in the reflection principle is always defined. Indeed,

$\{u: Pr(X \mid R(u)) = c\} \cap R(w)$ will often be empty, for example when $X = R(u)$ and $c < 1$. In a setting in which all evidential probabilities are defined, we treat the reflection principle as holding iff the above equation is satisfied whenever the outer probability is defined.

Other terminology: A frame $<W, R>$ is *quasi-reflexive* iff whenever wRx, xRx; $<W, R>$ is *quasi-symmetric* iff whenever wRx and xRy, yRx. Other frame conditions are as usual. In terms of a justified belief operator J with the usual accessibility semantics, quasi-reflexivity to the axiom $J(Jp \rightarrow p)$ and quasi-symmetry to the axiom $J(p \rightarrow J\neg J\neg p)$, and seriality to the axiom $Jp \rightarrow \neg J\neg p$.

Proposition 1. The reflection principle holds for a regular prior probability distribution over a serial frame only if the frame is quasi-reflexive, quasi-symmetric and transitive.

Proof: Suppose that the reflection principle holds for a regular probability distribution Pr over a serial frame $<W, R>$, and wRx. Since $<W, R>$ is serial and Pr regular, all evidential probabilities are defined.

(1) For quasi-reflexiveness, we show that xRx. Let $Pr(\{x\} \mid R(x)) = c$. Thus $x \in \{u: Pr(\{x\} \mid R(u)) = c\} \cap R(w)$ since wRx, so $Pr(\{x\} \mid \{u: Pr(\{x\} \mid R(u)) = c\} \cap R(w))$ is defined by regularity. Hence by reflection:

$$Pr(\{x\} \mid \{u: Pr(\{x\} \mid R(u)) = c\} \cap R(w)) = c$$

Therefore, since $x \in \{u: Pr(\{x\} \mid R(u)) = 0\} \cap R(w)$, $c > 0$ by regularity. Hence $Pr(\{x\} \mid R(x)) > 0$, so $\{x\} \cap R(x) \neq \{\}$, so xRx.

(2) For transitivity, we suppose that *xRy* and show that *wRy*. By regularity, $Pr(\{y\} \mid R(x)) = b > 0$. Hence $x \in \{u: Pr(\{y\} \mid R(u)) = b\} \cap R(w)$, so by regularity $Pr(\{y\} \mid \{u: Pr(\{y\} \mid R(u)) = b\} \cap R(w))$ is defined, so by reflection

$$Pr(\{y\} \mid \{u: Pr(\{y\} \mid R(u)) = b\} \cap R(w)) = b > 0$$

Therefore $\{y\} \cap R(w) \neq \{\}$, so *wRy*.

(3) For quasi-symmetry, we suppose that *xRy* and show that *yRx*. By quasi-reflexiveness, $y \in R(x) \cap R(y)$, *so* by regularity $Pr(R(y) \mid R(x)) = a > 0$. But by quasi-reflexiveness again $x \in \{u: Pr(R(y) \mid R(u)) = a\} \cap R(x))$, so $Pr(R(y) \mid \{u: Pr(R(y) \mid R(u)) = a\} \cap R(x))$ is defined by regularity, so by reflection

$$Pr(R(y) \mid \{u: Pr(R(y) \mid R(u)) = a\} \cap R(x)) = a$$

Suppose that $a < 1$. Whenever $u \in R(y)$, $R(u) \subseteq R(y)$ by transitivity, so $Pr(R(y) \mid R(u)) = 1$; thus $R(y) \cap \{u: Pr(R(y) \mid R(u)) = a\} = \{\}$, so $Pr(R(y) \mid \{u: Pr(R(y) \mid R(u)) = a\} \cap R(x)) = 0$, so $a = 0$, which is a contradiction. Hence $a = 1$. Thus $Pr(R(y) \mid R(x)) = 1$, so $R(x) \subseteq R(y)$ by regularity. But $x \in R(x)$, so $x \in R(y)$, so *yRx*.

Corollary 2. The reflection principle holds for a regular prior probability distribution over a reflexive frame only if the frame is partitional.

Proof: By Proposition 1; any reflexive quasi-symmetric relation is serial and symmetric.

Proposition 3. The reflection principle holds for any probability distribution over any finite quasi-reflexive quasi-symmetric transitive frame for which all evidential probabilities are defined.

Proof: Let *Pr* be a probability distribution over a finite quasi-reflexive quasi-symmetric transitive frame <*W, R*>. Pick $w \in W$, $X \subseteq W$ and a real number *c*. Suppose that $Pr(\{u: Pr(X \mid R(u)) = c\} \cap R(w)) > 0$, so $\{u: Pr(X \mid R(u)) = c\} \cap R(w) \neq \{\}$. Since *R* is quasi-reflexive, quasi-symmetric and transitive it partitions *R(w)*. Suppose that $x \in R(w)$. By transitivity, $R(x) \subseteq R(w)$. Moreover, if $y \in R(x)$ then by quasi-symmetry and transitivity $R(y) = R(x)$, so $Pr(X \mid R(y)) = Pr(X \mid R(x))$, so if $Pr(X \mid R(x)) = c$ then $Pr(X \mid R(y)) = c$. Hence if $x \in \{u: Pr(X \mid R(u)) = c\} \cap R(w)$ then $R(x) \subseteq \{u: Pr(X \mid R(u)) = c\} \cap R(w)$. Thus for some finite nonempty $Y \subseteq \{u: Pr(X \mid R(u)) = c\} \cap R(w)$:

$$\{u: Pr(X \mid R(u)) = c\} \cap R(w) = \cup_{y \in Y} R(y)$$

where $R(y) \cap R(z) = \{\}$ whenever *y* and *z* are distinct members of *Y*. Trivially, if $y \in Y$ then $Pr(X \mid R(y)) = c$. By hypothesis, *Pr(R(y))* is always positive. Consequently:

$$\begin{aligned} Pr(X \mid \{u: Pr(X \mid R(u)) = c\} \cap R(w)) &= Pr(X \mid \cup_{y \in Y} R(y)) \\ &= \textstyle\sum_{z \in Y} Pr(X \mid R(z)).Pr(R(z) \mid \cup_{y \in Y} R(y)) \\ &= \textstyle\sum_{z \in Y} cPr(R(z) \mid \cup_{y \in Y} R(y)) \\ &= c\textstyle\sum_{z \in Y} Pr(R(z) \mid \cup_{y \in Y} R(y)) \\ &= cPr(\cup_{z \in Y} R(z) \mid \cup_{y \in Y} R(y)) \\ &= c \end{aligned}$$

Proposition 4: The reflection principle holds for any countably additive probability distribution over any quasi-reflexive quasi-symmetric transitive frame for which all evidential probabilities are defined.

Proof: The proof is like that for Proposition 3. In particular, whatever the cardinality of *W*, $\{R(y)\}_{y \in Y}$ is a family of disjoint sets to each of which *Pr* gives positive probability, so *Y* must be at most countably infinite by a familiar property of real-valued distributions. Thus the summation in the proof is over a countable set.

Corollary 5. For a regular probability distribution over a finite serial frame, the reflection principle holds iff the frame is quasi-reflexive, quasi-symmetric and transitive.

Proof: From Propositions 1 and 3, since all evidential probabilities are defined for a regular probability distribution over a serial frame.

Corollary 6. For a regular probability distribution over a finite reflexive frame, the reflection principle holds iff the frame is partitional.

Proof: From Propositions 2 and 3.

Corollary 7. For a regular countably additive probability distribution over a serial frame, the reflection principle holds iff the frame is quasi-reflexive, quasi-symmetric and transitive.

Proof: From Propositions 1 and 4.

Corollary 8. For a regular countably additive probability distribution over a reflexive frame, the reflection principle holds iff the frame is partitional.

Proof: From Propositions 2 and 4.

Notes

[*] This paper developed out of Williamson 2011. The appendix was prompted by discussion with Horacio Arlo-Costa, and section 5 by discussion with Kevin Kelly and Hanti Lin. I have also added some brief comments in response to Conee and Feldman 2011. The idea for the original paper arose in response to discussion with John Hawthorne. Earlier versions of the material were presented at the following venues: classes and a discussion groups at Oxford, an Arché workshop on Basic Knowledge at St Andrews (Philip Ebert was the commentator), a workshop on epistemology at the University of Wisconsin at Madison (Jim Pryor was the commentator), graduate conferences at University College London and the University of Miami, a CSMN/Arché graduate conference at the University of Oslo, a workshop on mathematical methods in philosophy in Banff (Alberta), a symposium for Jaakko Hintikka in Copenhagen, a workshop on epistemic aspects of many-valued logic at the Institute of Philosophy of the Academy of Sciences of the Czech Republic in Prague, a conference of the Italian Society for Analytic Philosophy in Padua, the Moral Sciences Club in Cambridge, the Argentine Society for Analytic Philosophy in Buenos Aires, the Serbian Philosophical Society in Belgrade, colloquia in the philosophy departments of Lund University, Nottingham University, the University of Texas at Austin, Brandeis University, Indiana University at Bloomington, McGill University, and Carnegie Mellon University. I thank all those who have contributed to the paper with their comments and questions.

[1] For an argument that epistemically possible propositions can have probability 0, even when infinitesimal probabilities are allowed, see Williamson 2007b.

[2] The talk of discrimination is just shorthand for talk of how much the subject can know about what time it is (for more discussion see Williamson 1990). In their argument, Conee and Feldman 2011 assume that indiscriminable positions of the hand appear visually to the subject in the same way, but this assumption is unwarranted; they might appear in different ways between which the subject cannot discriminate.

[3] Concerning an almost exactly similar example, although it was set up in terms of position rather than time, Conee and Feldman 2011 claim both that ‘Even the known proposition is not stated’ and that ‘The proposition that the pointer points somewhere in the relevant arc is the proposition that S [the subject] knows’. Their confusion may result from the presentation of the known proposition as *R(w)*, rather than by a sentence you, the subject, use to express it. They then suppose that the relevant sentence is ‘It is pointing somewhere in there’, where S means ‘there’ to identify ‘S’s exact discriminatory limits as S knows them to be’. As explained in the text, this is not the pertinent way to take the example.

[4] Unlike the alternative examples of improbable knowing proposed by Conee and Feldman 2011, the example does not depend on a belief condition on knowledge, but rather on its more specifically epistemic character. This is important for its role in

explaining why we may be reluctant to ascribe knowledge of *p* to someone who on their own evidence is unlikely to know *p*.

5 One should not get the impression that the case against the KK principle itself depends on the use of standard formal models of epistemic models. The anti-KK argument at Williamson 2000: 114-18 makes no such appeal. Their use here is to enable the calculation of evidential probabilities.

6 On an epistemic account of vagueness, such variable margins for error yield distinctive forms of higher-order vagueness. Williamson (1999: 136-8) argues that if the ‘clearly’ operator for vagueness obeys the analogue of the B (for ‘Brouwersche’) axiom $p \to K\neg K\neg p$ (which corresponds to the condition of symmetry on *R*) then any formula with second-order vagueness has *n*th-order vagueness for every $n > 2$, but does not endorse the B axiom for ‘clearly’. In response, Mahtani (2008) uses variable margins for error to argue against the B axiom for ‘clearly’ and suggests that they allow vagueness to cut out at any order. Dorr (2008) provides a formal model in which he proves Mahtani’s suggestion to hold. These arguments all have analogues for the overtly epistemic case.

7 Stalnaker 2006 has the axiom schema $Bp \to BKp$ (which entails $Bp \to BK^n p$ for arbitrary iterations K^n of *K*).

8 For a recent discussion of the relation between knowledge and reasons for action see Hawthorne and Stanley 2008; see also Hyman 1999 and Gibbons 2001.

[9] For a critique of internalist misinterpretations of excuses as justification see Williamson 2007a.

[10] Such a principle is called ‘intuitive closure’ at Williamson 2000: 117-18.

[11] For discussion of MPC in relation to the Preface Paradox see Hawthorne 2004: 46-50, 154, 182-3. Similar cases involving future chances are used in Hawthorne and Lasonen-Aarnio 2009 against the safety conception of knowledge in Williamson 2000; for a reply see Williamson 2009.

[12] A similar problem arises for single-premise closure principles when one competently carries out a long chain of single-premise deductive steps, each with a small epistemic probability of inferential error (in the multi-premise case, for simplicity, one’s deductive competence is treated as beyond doubt); see Lasonen-Aarnio 2008 for discussion. Here is a parallel account of that case. One knows the premise without knowing that one knows it. For each deductive step, one carries it out competently without knowing that one does so. By single-premise closure, one knows the conclusion, without knowing that one knows it. For each deductive step, it is very probable on one’s evidence that one carries it out competently. However, it is very improbable one one’s evidence that one carries out every deductive step competently. Since it is granted that one knows the conclusion only if one carries out every deductive step competently, it is very improbable on one’s evidence that one knows the conclusion.

[13] The usual form of epistemic modelling is not appropriate for treating possible errors in deductive reasoning, since logical omniscience suppresses the dependence of inferential knowledge on correct inferential processes.

[14] The technical details are taken from Williamson 2009.

[15] Proof: Suppose that $w_i > k$. If wRx then $|w_i – x_i| \leq k$, so $x_i > 0$, so $x \in p_i$. Thus $w \in Kp_i$. Conversely, suppose that $w_i \leq k$. Then $wRw[i|0]$, for $|w_i – w[i|0]_i| = |w_i – 0| = w_i \leq k$ and if $i \neq j$ then $|w_j – w[i|0]_j| = 0$; but $w[i|0] \notin p_i$ because $w[i|0]_i = 0$, so $w \notin Kp_i$.

[16] Proof: Suppose that q is i-based and $x_i = y_i$. Suppose also that $x \notin Kq$. Then for some z, xRz and $z \notin q$. But then $yRy[i|z_i]$, for $|y_i – y[i|z_i]_i| = |y_i – z_i| = |x_i – z_i|$ (because $x_i = y_i$) $\leq k$ (because xRz), and if $i \neq j$ then $|y_j – y[i|z_i]_j| = 0$. Moreover, $y[i|z_i] \notin q$ because $z \notin q$, q is i-based and $y[i|z_i]_i = z_i$. Hence $y \notin Kq$. Thus if $y \in Kq$ then $x \in Kq$. By parity of reasoning the converse holds too.

[17] Proof: Set $\#(i, q, w) = \{j\colon 0 \leq j \leq 2k,\ w[i|j] \in q \text{ and } |w_i – j| \leq k\}$ for any $w \in W$, $q \subseteq W$, $1 \leq i \leq n$. For each i, let q_i be i-based. Let $\cap q_i = q_1 \cap \ldots \cap q_n$. For $w \in W$, $R(w) \cap \cap q_i = \{x\colon \forall\ i, x_i \in \#(i, q_i, w)\}$, since for each i and $x \in W$, $x \in q_i$ iff $w[i|x_i] \in q_i$ since q_i is i-based. Since $\text{Prob}_{\text{prior}}$ is uniform, $\text{Prob}_w(\cap q_i) = |\{R(w) \cap \cap q_i|/|R(w)|$ for $w \in W$. But $|R(w) \cap \cap q_i| = |\{x\colon \forall\ i, x_i \in \#(i, q_i, w)\}| = |\#(1, q_1, w)| \ldots |\#(n, q_n, w)|$. By the special case

of this equation in which each q_i is replaced by W (which is trivially i-based for any i), $|R(w)| = |\#(1, W, w)|\ldots |\#(n, W, w)|$. Consequently:

$\mathrm{Prob}_w(\cap q_i) = (|\#(1, q_1, w)|\ldots |\#(n, q_n, w)|)/(|\#(1, W, w)|\ldots |\#(n, W, w)|)$.

For any given i, consider another special case in which q_j is replaced by W whenever $i \neq j$. Since n–1 of the ratios cancel out, $\mathrm{Pr}_w(q_i) = |\#(i, q_i, w)|/|\#(i, W, w)|$. Therefore $\mathrm{Prob}_w(\cap q_i) = \mathrm{Prob}_w(q_1)\ldots \mathrm{Prob}_w(q_n)$, as required.

18 Proof: We have already established that $x \in Kp_i$ iff $x_i > k$. Thus, in the notation of the previous footnote, $\#(i, Kp_i, <2k, \ldots, 2k>) = \{j: k < j \leq 2k\}$, so $|\#(i, Kp_i, <2k, \ldots, 2k>)| = k$, while $\#(i, W, <2k, \ldots, 2k>) = \{j: k \leq j \leq 2k\}$, so $|\#(i, W, <2k, \ldots, 2k>)| = k+1$. By the formula for $\mathrm{Prob}_w(q_i)$ in the previous footnote (with Kp_i in place of q_i), $\mathrm{Prob}_{<2k,\ldots,2k>}(Kp_i) = k/(k+1)$.

19 A similar generalization to higher iterations of knowledge is possible for the case of multiple risks of inferential error in a single-premise deduction. One has at least n iterations of knowledge of the premise. For each deductive step, one has $n-1$ but not n iterations of knowledge that one carried it out competently. By single-premise closure and plausible background assumptions, one has n but not n+1 iterations of knowledge of the conclusion. For each deductive step, it is very probable on one's evidence that one has at least $n-1$ iterations of knowledge that one carried it out competently. However, it is very improbable one one's evidence that one has at least $n-1$ iterations of knowledge that one carried out every deductive step competently. Since it is granted that one has at least n iterations of knowledge of the conclusion only if one has at least $n-1$ iterations of

knowledge that one carried out every deductive step competently, it is very improbable on one's evidence that one has at least *n* iterations of knowledge of the conclusion.

[20] See Williamson 2008 for more discussion of the structure and semantics of higher-order evidential probabilities. The phenomenon discussed in the text involves the apparent loss of only one iteration of knowledge between premises and conclusion. However, the apparent absence of a given number of iterations of knowledge can cause doubts about all lower numbers of iterations, by a domino effect, since lack of knowledge that one has *n*+1 iterations implies lack of warrant to assert that one has *n* iterations (Williamson 2005: 233-4).

Bibliography

Conee, E., and Feldman, R. 2011. 'Response to Williamson', in Dougherty 2011.

Dorr, C. 2008. 'How vagueness could cut out at any order'. MS.

Dougherty, T., ed. 2011. *Evidentialism and its Discontents*. Oxford: Oxford University Press.

Gettier, E. 1963. 'Is justified true belief knowledge?'. *Analysis* 23: 121-3.

Gibbons, J. 2001. 'Knowledge in action', *Philosophy and Phenomenological Research* 62: 579-600.

Goldman, A. 1976. 'Discrimination and perceptual knowledge'. *The Journal of Philosophy* 73: 771-91.

Greenough, P., and Pritchard, D., eds. 2009. *Williamson on Knowledge*, Oxford: Oxford University Press.

Hawthorne, J. 2004. *Knowledge and Lotteries*. Oxford: Clarendon Press.

Hawthorne, J., and Lasonen-Aarnio, M. 2009. 'Knowledge and objective chance', in Greenough and Pritchard.

Hawthorne, J., and Stanley, J. 2008. 'Knowledge and action'. *The Journal of Philosophy* 105: 571-90.

Hintikka, J. 1962. *Knowledge and Belief*. Ithaca, N.Y.: Cornell University Press.

Hyman, J. 1999. 'How knowledge works'. *The Philosophical Quarterly* 49: 433-51.

Lasonen-Aarnio, M. 2008. 'Single premise deduction and risk'. *Philosophical Studies* 141: 157-73.

Lemmon, E. J. 1967. 'If I know, do I know that I know?'. In A. Stroll (ed.), *Epistemology*. New York: Harper & Row.

Mahtani, A. 2008. 'Can vagueness cut out at any order?'. *Australasian Journal of Philosophy* 86: 499-508.

Radford, C. 1966. 'Knowledge — by examples'. *Analysis* 27: 1-11.

Stalnaker, R. 1999. *Context and Content: Essays on Intentionality in Speech and Thought*. Oxford: Oxford University Press.

Stalnaker, R. 2006. 'On logics of knowledge and belief'. *Philosophical Studies* 128: 169 -99.

Williamson, T. 1990. *Identity and Discrimination*. Oxford: Blackwell.

Williamson, T. 1999. 'On the structure of higher-order vagueness'. *Mind* 108: 127-43.

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

Williamson, T. 2005. 'Contextualism, subject-sensitive invariantism and knowledge of knowledge'. *The Philosophical Quarterly* 55: 213-35.

Williamson, T. 2007a. 'On being justified in one's head'. In M. Timmons, J. Greco and A. Mele (eds.), *Rationality and the Good: Critical Essays on the Ethics and Epistemology of Robert Audi*. Oxford: Oxford University Press.

Williamson, T. 2007b. 'How probable is an infinite sequence of heads?' *Analysis* 67: 173-80.

Williamson, T. 2008. 'Why epistemology can't be operationalized.' In Q. Smith (ed.), *Epistemology: New Philosophical Essays*. Oxford: Oxford University Press.

Williamson, T. 2009. 'Reply to John Hawthorne and Maria Lasonen-Aarnio', in Greenough and Pritchard.

Williamson, T. 2011. 'Improbable knowing'. In Dougherty 2011.