PROBABILITY AND DANGER

TIMOTHY WILLIAMSON

THE AMHERST LECTURE IN PHILOSOPHY

LECTURE 4, 2009

http://www.amherstlecture.org/

Probability and Danger

Timothy Williamson

PREFERRED CITATION

Williamson, Timothy. "Probability and Danger." *The Amherst Lecture in Philosophy* 4 (2009): 1–35. <http://www.amherstlecture.org/williamson2009/>.

ABSTRACT

What is the epistemological structure of situations where many small risks amount to a large one? Lottery and preface paradoxes and puzzles about quantum-mechanical blips threaten the idea that competent deduction is a way of extending our knowledge (MPC). Seemingly, everyday knowledge involves small risks, and competently deducing the conjunction of many such truths from them yields a conclusion too risky to constitute knowledge. But the dilemma between scepticism and abandoning MPC is false. In extreme cases, objectively improbable truths are known. Safety is modal, not probabilistic, in structure, with closure and factiveness conditions. It is modelled using closeness of worlds. Safety is analogous to knowledge. It suggests an interpretation of possible worlds semantics for epistemic logic. To avoid logical omniscience, a relation of epistemic counterparthood between formulas is introduced. This supports a safety conception of knowledge and formalizes how extending knowledge by deduction depends on logical competence.

The Amherst Lecture in Philosophy (ISSN: 1559-7199) is a free on-line journal, published by the Department of Philosophy, Amherst College, Amherst, MA 01002. Phone: (413) 542-5805. E-mail: alp@amherst.edu. Website: http://www.amherstlecture.org/.

Probability and Danger

Timothy Williamson

University of Oxford

1. MUCH OF RECENT and not-so-recent philosophy is driven by tensions, or at least apparent tensions, between common sense and natural science – in the terminology of Wilfrid Sellars, between the manifest image and the scientific image of the world. These tensions arise most saliently in metaphysics and the philosophy of mind, but are far from confined to those branches of philosophy. In this lecture, I will discuss one specific form they take in contemporary epistemology.

Central to common sense epistemology is the distinction between knowledge and ignorance. Knowledge is not usually conceived as coming in quantifiable degrees: we do not ask and could not answer 'To what degree does she know where the station is?'.[1] By contrast, a continuum of numerical degrees of probability is central to contemporary natural science. The point is not merely that a framework of probabilities has to some extent displaced a framework of knowledge and ignorance in the scientific image of cognition. Worse, probabilistic reasoning seems to destabilize common sense conceptions of knowledge. As so often, we cannot just blandly assert that the manifest image and the scientific image are both fine in their own way, but useful for different purposes. We face *prima facie* conflicts between them which seem to imply that if the scientific image is accurate, then the manifest image is radically misleading. We have to do the hard work of analysing the apparent conflicts in detail, to determine what their upshot really is.

1 For discussion of the (un)gradability of knowledge ascriptions see Stanley 2005: 35–46.

Elsewhere, I have argued that the sort of probability most relevant to the epistemology of science is probability on the evidence, and that the evidence is simply what is known; thus knowledge is a precondition, not an outdated rival, of probability in science (Williamson 2000). I have also shown that much of the supporting argument for that conclusion is robust even when recast in probabilistic form (Williamson 2008). But those arguments do not entitle us to ignore specific probabilistic considerations that seem to undermine common sense epistemology. This lecture concerns one such threat to a non-probabilistic conception of knowledge.

2. WHY IS DEDUCTION USEFUL? The obvious answer is that it is a way of extending our knowledge. It is integral to that answer that extending one's knowledge in this way depends on the temporal process of carrying out the deduction, for one knows more after doing so than one did before. Moreover, one cannot expect to obtain knowledge thereby unless the deductive process involves forming a belief in the conclusion. This suggests a principle along the following lines, now often known as 'Multi-Premise Closure' (Williamson 2000: 117):

> MPC If one believes a conclusion by competent deduction from some premises one knows, one knows the conclusion.

Here competence is intended to stand to inference roughly as knowledge stands to belief. One can no more hope to attain knowledge of the conclusion by less than competent deduction than one can hope to attain it by deduction from premises of which one has less than knowledge. But competence does not require knowledge that the deduction is valid, otherwise the attempt to use MPC to explain how deduction extends knowledge would involve an infinite regress of knowledge of the validity of more and more complex deductions (Carroll 1895). MPC is closer to the dynamics of cognition than is the static principle that one knows a conclusion if one believes it, knows it to follow deductively from some premises, and knows the premises.

At first sight, there is no tension between MPC and a scientific account of cognition. Mathematics is essential to science, and its main role is to extend our knowledge by deduction.

Perhaps some fine-tuning is needed to capture exactly the intended spirit of MPC. Nevertheless, some such principle seems to articulate the compelling idea that deduction is a way of extending our knowledge. I will not discuss any fine-tuning of MPC here. Nor will I discuss challenges to MPC that are closely related to traditional sceptical puzzles, for instance where the premise is 'That is a zebra' and the conclusion is 'That is not just a mule cleverly painted to look like a zebra'. It is generally, although not universally, agreed that such examples do not refute a properly formulated closure principle for knowledge.[2] Even if we start by answering the question 'Do the spectators know that that is a zebra?' in the affirmative and then the question 'Do they know that it is not just a mule cleverly painted to look like a zebra?' in the negative, once that has happened and we are asked again 'Do they know that it is a zebra?' we are now inclined to answer in the negative. Thus the supposed counter-example to closure is not stable under reflection.

The probabilistic threat to MPC starts from the truism that many acceptably small risks of error can add up to an unacceptably large one. The most obvious illustration is a version of the Lottery Paradox (Kyburg 1961). Suppose that for some positive real number δ a risk of error less than δ is acceptable. Then for any suitably large natural number n, in a fair lottery with n tickets of which only one wins, for each losing ticket the statement that it will lose has an acceptably small risk of error, but all those statements together logically entail their conjunction, which has a probability of only $1/n$ – the structure of the lottery being given – and *a fortiori* an unacceptably large risk of error. This does not constitute a clear counter-example to MPC, since one can deny that the premises are known: even if a ticket will in fact lose, we do not know in advance that it will; we only know that it is almost certain to. But can we legitimately treat knowledge of lotteries as a special case?[3] For example, does not a scientific study of human perception and memory show that even in the best cases they too involve non-zero risks of error? If we reacted to the Lottery Paradox by insisting that knowledge requires zero risk of error, that requirement seems to constrain us to denying that there is human knowledge by perception or memory, and more generally to force us into scepticism.

2 The example is of course from Dretske 1970; the other classic version of such a challenge to closure is Nozick 1981. For critical discussion of such objections to closure see Vogel 1990 and Hawthorne 2005.

3 See Hawthorne 2004 for discussion.

Even beliefs about our own present mental states seem to carry some non-zero risk of error. But if knowledge of our contingent circumstances is unobtainable, the distinction between knowledge and ignorance loses most of its interest.

A version of the Preface Paradox helps make the point vivid (Makinson 1965). Suppose that I compile a reference book containing large quantities of miscellaneous information. I take great care, and fortunately make not a single error. Indeed, by ordinary standards I know each individual item of information in the book. Still, I can reasonably acknowledge in the preface that since almost all such works contain errors, it is almost certain that mine does too. If I nevertheless believe the conjunction of all the individual items of information in the book (perhaps excluding the preface), the risk of error in that conjunctive belief seems so high that it is difficult to conceive it as knowledge. Thus MPC seems to fail, unless the standard for knowing is raised to sceptical heights.

One advantage of the objection to MPC from the Preface Paradox over the generalization from the Lottery Paradox is that it avoids the unargued assumption that if a true belief that a given ticket will lose fails to constitute knowledge, the reason must be just that it has a non-zero risk of error. For whether a true belief constitutes knowledge might depend on all sorts of factors beyond its risk of error: for example, its causal relations. By contrast, the objection from the Preface Paradox makes trouble simply by conjoining many miscellaneous items of what by common sense standards is knowledge; it does not depend on the subject matter of that putative knowledge.

The common sense epistemologist seems to face a dilemma: either reject MPC or become a sceptic. The first horn is not much better than the second for common sense epistemology. If deduction can fail to extend knowledge, through the accumulation of small risks, then an explicitly probabilistic approach seems called for, in order to take account of those small risks, and the distinction between knowledge and ignorance is again sidelined, just as it is on the sceptical horn.

However, the argument for the dilemma is less clear than it seems. It trades on an unexamined notion of risk. It treats risk as a probabilistic matter, but what sort of probability is supposed to be at issue? The problem does not primarily concern the agent's subjective probabilities (degrees of belief), for even if the agent has perfect confidence in every conjunct and their conjunction, that does not address the worry that the risk of error in the conjunction is too high for the agent's true belief in it to constitute knowledge. Nor do probabilities on the agent's evidence do the trick. For since the probability of any item of evidence on the evidence of which it is part is automatically 1, the probability of any conjunction of such items of evidence on that evidence is also 1. But whatever exactly the items of evidence are, some variant on the Preface Paradox will arise for them too. This may suggest that risk should be understood as a matter of objective probabilities (chances), at least for purposes of the argument.

In a recent paper, John Hawthorne and Maria Lasonen-Aarnio have developed just such a chance-based argument.[4] It can be adapted for present purposes as follows. Assume, with common sense, that we have at least some knowledge of the future. For example, I know that my carpet will remain on my floor for the next second. Nevertheless, as an instance of quantum indeterminacy, there is a non-zero chance, albeit a very small one, that the carpet will not remain on the floor for the next second, but will instead rise up into the air or filter through the floor. Now suppose that there are n carpets, each in a situation exactly like mine. Let p_i be the proposition that the ith carpet remains on the floor for the next second (for expository purposes, I write as though from a fixed time). Suppose also that nothing untoward will in fact happen, so all those propositions about the future are true:

(1) $p_1, \ldots, p_n$ are true.

We may assume:

(2) Each of $p_1, \ldots, p_n$ has the same high chance less than 1.

We may also assume, at least to a good enough approximation, that the carpets and their circumstances do not interact in any way that would make the chances of some of them remaining on the floor depend on the chances of others doing so; the n propositions about the future are independent of each other in the sense that the chance of any conjunction of them is simply the product of the chances of the conjuncts. In brief:

(3) $p_1, \ldots, p_n$ are mutually probabilistically independent.

4 See Hawthorne and Lasonen-Aarnio 2009. I have omitted various subtleties from the argument that are not of present concern; my reply in Williamson 2009 pays more attention to them.

For large enough n, (1), (2) and (3) together entail (4):

(4) The chance of p_1 & … & p_n is low.

Imagine that one is in a good position to monitor each carpet. One believes of each carpet that it will remain on the floor, competently deduces the conjunction of all those propositions from the conjuncts and thereby comes to believe that all the carpets will remain on the floor:

(5) One believes p_1 & … & p_n by competent deduction from $p_1, \ldots, p_n$.

We may treat (1)–(5) as a relevantly uncontentious description of the example. A further plausible claim about the example is that one knows of each carpet that it will remain on the floor, in much the way that I do with my carpet:

(6) One knows each of $p_1, \ldots, p_n$.

A plausible general constraint is that one cannot know something unless it has a high chance of being true:

HC One knows something only if it has a high chance.

Unfortunately, relative to the uncontentious description of the example (1)–(5), (6) forms an inconsistent triad with MPC and HC. For (5), (6) and MPC together entail (7):

(7) One knows p_1 & … & p_n.

But (4) and HC together entail the negation of (7):

(8) One does not know $p1$ & … & p_n.

Holding (1)–(5) fixed, we must give up at least one of (6), MPC and HC. Which should it be?

Although giving up (6) permits us to retain both MPC and HC, it amounts to scepticism, at least with respect to knowledge of the future. If we remain anti-sceptics and retain HC, we must give up MPC, and face the consequent problems. Later, I will assess the third option, giving up HC in order to combine anti-scepticism with MPC, and argue that it is much more natural than it sounds. Before doing so, however, I will explore some probabilistic aspects of the argument in more detail.

3. ACCORDING TO SOME PHILOSOPHERS, the principle of bivalence fails for future contingents. On their view, p_i is neither true nor false in advance, because the chancy future is a mere range of possibilities until one of them comes to pass. Thus examples of the kind supposed are impossible, because (1) is incompatible with (2): p_i cannot be simultaneously true and chancy. Presumably, on this view, since p_i cannot be true in advance, it also cannot be known in advance. Thus (6) is denied, as well as (1). This is a form of scepticism with respect to knowledge of the future, but its motivation is quite specific; it does not threaten to spread into a more general scepticism. Truths about the past are supposed to have chance 1. This view of the future makes the argument uninteresting. However, since I accept bivalence for future contingents – they are true or false in advance, whether or not we can already know which – I will not try to defuse the argument in that way. I accept (1)–(5) as a description of a genuine possibility.

Nevertheless, the resort to objective chance in the argument is curious. For, on the face of it, the problem does not depend on objective chance. For example, suppose that the observer is isolated for the relevant second, unable to receive new perceptual information about the fate of the carpets. At the end of that second, the relevant propositions have become straightforward truths about the past. But the same epistemological problem seems to arise: belief in the conjunction p_1 & … & p_n still seems too risky to constitute knowledge, even though the risk is not a matter of objective chance. More generally, the Preface Paradox seems to raise the same problem, irrespective of the specific subject-matter of the conjoined propositions. Even if our universe turns out to be deterministic and devoid of objective chance, we still face situations in which many acceptably small risks of error in the premises accumulate into an unacceptably large risk of error in the conclusion. Although posing the problem in terms of objective chances makes it especially vivid, it also makes its underlying nature harder to discern. Perhaps we need a kind of probability that is less objective than objective chance, but more objective than probability on the evidence, in order to capture the relevant notion of risk.

Whatever the relevant kind of probability, it should obey the standard axioms of the probability calculus, and we can make some points on that basis. The starting point for the problem is that if δ is any positive real number not greater than 1, there are deductively valid arguments each of whose premises has a probability greater than $1-\delta$ but whose conclusion has a probability not greater than $1-\delta$. Any such argument has at least two premises. For the probability axioms guarantee that when a conclusion follows deductively from a single premise, the probability of the conclusion is at least as high as the probability of the premise, and when a conclusion follows deductively from the null set of premises, the probability of the conclusion is 1 (because it is a logical truth). The problem therefore seems to be essentially one for *multi*-premise closure (MPC), and not to arise for single-premise closure:

SPC If one believes a conclusion by competent deduction from a premise one knows, one knows the conclusion.

On further reflection, however, that is not a satisfying result. For what seems to be a version of the same problem arises for single-premise closure too (Lasonen-Aarnio 2008). The reason is that the process of deduction involves its own risks. We are no more immune from errors of logic than we are from any other sort of error. The longer a chain of reasoning extends, the more likely it is to contain mistakes. One might even know that in the past one made on average about one mistake per hundred steps of reasoning, so that a chain of one's reasoning a thousand steps long is almost certain to contain at least one mistake. Suppose that one knows *p*, and does in fact carry out each step of the long deduction competently, thereby eventually arriving at a belief in *q*. By repeated applications of SPC, one knows each intermediate conclusion and finally *q* itself (surely carrying out the later steps does not make one cease to know the earlier conclusions). But the same worry as before about the accumulation of many small risks of error still arises.

Of course, we can still ask probabilistic questions about the process of deduction itself. For example, what is the probability that the conclusion of an attempted deduction by me is true, conditional on the assumption that the premise is true and the attempted deduction contains a thousand steps, each with an independent probability of 1/100 of containing a mistake? The difficulty is to know which probabilistic questions to ask. The question just formulated abstracts from the identity of the premise and conclusion of the attempted deduction, but not from its length. Why should that be the relevant abstraction? After all, the reasoner usually knows the identity of the premises and conclusion quite well. If we specify their identity in the assumption, and the attempted deduction is in fact valid, then the conditional probability is 1 again, and the risk seems to have disappeared. This, of course, is an instance of the notorious reference class problem, which afflicts many theories of probability.[5] But it is of an especially pressing form, because the apparent risk can only be captured probabilistically by abstracting from intrinsic features of the deduction and subsuming it under a general reference class.

None of this shows that probability does not have an essential role to play in the understanding of risk; surely it has. However, when probabilities are invoked, much of the hardest work will consist in the prior analysis of the issues that explains why one reference class rather than another is relevant. When epistemological risk is at issue, the explanation will be in epistemological terms; invoking probabilities to explain why a given reference class is the relevant one would merely postpone the problem.

4. IN THINKING ABOUT the epistemological problem of risk, it is fruitful to start from a conception of knowledge as safety from error. I have developed and defended such a conception elsewhere.[6] I do not intend it to provide necessary and sufficient conditions for knowing in more basic terms. Without reference to knowing, it would be too unclear what sort of safety was in question. Rather, the point of the safety slogan is to suggest an analogy with other sorts of safety that is useful in identifying some structural features of knowledge.

For comparison, think of David Lewis's similarity semantics for counterfactual conditionals. Its value is not to enable one to determine whether a counterfactual is true in a given case by applying one's general understanding of similarity to various possible worlds, without reference to counterfactuals themselves. If one tried to do that, one would almost certainly give the wrong comparative weights to the various relevant respects of similarity. Nevertheless, the semantics gives valuable structural information about counterfactuals, in particular about their logic. Likewise, the point of a safety conception of knowing is not to

5 See Hájek 2007 for a recent discussion.

6 See Williamson 1992 and 2000: 123–130. For a related notion of safety see Sosa: 1996, 2000 and 2007.

enable one to determine whether a knowledge attribution is true in a given case by applying one's general understanding of safety, without reference to knowing itself. If one tried to do that, one would very likely get it wrong. Nevertheless, the conception gives valuable structural information about knowing.[7] The considerations that follow are intended in that spirit.

There seem to be two salient rival ways of understanding safety in terms of risk. On the 'no risk' conception of safety, being safe from an eventuality consists in there being no risk of its obtaining. On the 'small risk' conception of safety, being safe from an eventuality consists in there being at most a small risk of its obtaining. The two conceptions disagree on whether a low but non-zero level of risk excludes or implies safety. Each conception of safety combines with a general conception of knowledge as safety from error to yield a more specific conception of knowledge. The safety conception of knowledge and a 'no risk' conception of safety jointly imply a 'no risk of error' conception of knowledge. The safety conception of knowledge and a 'small risk' conception of safety jointly yield a 'small risk of error' conception of knowledge.

At first sight, the 'no risk of error' conception of knowledge imposes an unrealistically high, infallibilist standard on human cognition that leads to scepticism, and in particular forces rejection of (6) while allowing retention of both MPC and HC. From the same perspective, the 'small risk of error' conception of knowledge seems to impose a more realistically low, fallibilist standard on human cognition that avoids scepticism, and in particular permits

7 See Lewis 1973: 91–5 and 1986: 52–5 for his conception of similarity. At Lewis 1986: 41 (reprinted from Lewis 1979) he writes: 'Analysis 2 (plus some simple observations about the formal character of comparative similarity) is about all that can be said in full generality about counterfactuals. While not devoid of testable content – it settles some questions of logic – it does little to predict the truth values of particular counterfactuals in particular contexts. The rest of the study of counterfactuals is not fully general. Analysis 2 is only a skeleton. It must be fleshed out with an account of the appropriate similarity relation, and this will differ from context to context.' Lewis then makes it clear that, in the latter task, we must use our judgments of counterfactuals to determine the appropriate similarity relation.

An example of structural information about knowledge that can be extracted from the safety conception is anti-luminosity: only trivial conditions obtain only when one is in a position to know that they obtain (Williamson 2000: 96–109). For replies along these lines to some critics of the safety conception see Williamson 2009.

retention of (6) as well as HC while forcing rejection of MPC. This makes the 'small risk of error' conception of knowledge look the more attractive of the two, even though its rejection of MPC is initially unpleasant and makes the usefulness of deduction harder to explain.

One immediate problem for the 'small risk of error' conception of knowledge is that, unrevised, it is incompatible with the factiveness of knowledge: only truths are known. If *p* is false, you don't know *p*, even if you believe that you know *p*. For if the risk of error is small but not nonexistent, error may still occur. Although one could revise the 'small risk of error' conception of knowledge by adding truth as an extra conjunct, such *ad hoc* repairs count against a theory.

In order to decide between the two safety conceptions of knowledge, it is useful to step back from epistemology and consider the choice between the corresponding conceptions of safety in general. By reflecting on the non-technical distinction between safety and danger, especially in non-epistemological settings where we have fewer theoretical preconceptions, we can see the epistemological issues from a new angle. After all, the distinction between safety and danger is not in general an epistemological one. For example, whether one is safe from being abducted by aliens is a completely different question from whether one knows or believes that one will not be abducted by aliens.

5. HERE ARE TWO ARGUMENTS ABOUT SAFETY that seem to be valid, when the context is held fixed between premises and conclusion:

Argument A_{safety}

S was shot.

S was not safe from being shot.

Argument B_{safety}

S was safe from being shot by X.

S was safe from being shot by Y.

S was safe from being shot by Z.

S was safe from being shot other than by X, Y or Z.

S was safe from being shot.

On a 'small risk' conception of safety, neither argument is valid. Indeed, the corresponding arguments explicitly about small risks do not even look particularly plausible:

Argument $A_{smallrisk}$

S was shot.

S's risk of being shot was not small.

Argument $B_{smallrisk}$

S's risk of being shot by X was small.

S's risk of being shot by Y was small.

S's risk of being shot by Z was small.

S's risk of being shot other than by X, Y or Z was small.

S's risk of being shot was small.

In the case of argument $A_{smallrisk}$, it is obvious that one may be shot even if one's risk of being shot is small. In the case of argument $B_{smallrisk}$, it is almost equally obvious that many small risks may add up to a large one.

By contrast, on a 'no risk' conception of safety, both arguments are valid.[8] The corresponding arguments explicitly about the absence of risks look compelling:

Argument A_{norisk}

S was shot.

S was at some risk of being shot.

8 Argument A_{norisk} would be invalid if 'no risk' were equated with 'zero risk' in a probabilistic sense, for events of probability zero can occur. This holds even if infinitesimal probabilities are allowed (Williamson 2007). In quantitative terms, the 'no risk' conception is not the limiting case of the 'small risk' conception. The 'zero risk' conception of safety is intermediate between the 'no risk' and 'small risk' conceptions; it does validate argument B_{safety}. I do not discuss it at length here because it is a hybrid that combines many of the features of the 'no risk' and 'small risk' conceptions that each side objects to in the other's view.

Argument B_{norisk}

S was at no risk of being shot by X.

S was at no risk of being shot by Y.

S was at no risk of being shot by Z.

S was at no risk of being shot other than by X, Y or Z.

S was at no risk of being shot.

These results strongly suggest that we ordinarily think of safety according to a 'no risk' rather than a 'small risk' conception. Given the ease with which we classify the 'small risk' arguments as invalid, it is quite implausible that we have a 'small risk' conception of safety but are unable to work out its consequences for A_{safety} and B_{safety}. But then, if the 'no risk of error' conception of knowledge commits us unwittingly to epistemological scepticism, presumably the 'no risk' conception of safety likewise commits us unwittingly to the analogue of scepticism for safety in general, the view that we are virtually never safe from anything. Alternatively, in order to avoid such generalized scepticism, it might be held that, contrary to appearances, arguments A_{safety} and B_{safety} are in fact invalid. Either way, we turn out to be radically mistaken about the nature of safety, in one case about its structure, in the other about its extent. Nor is any explanation of our radical misconceptions in the offing.

Those are unattractive hypotheses. Fortunately, we are not forced to choose between them, for we have been considering too narrow a range of theoretical options. An alternative is to retain the 'no risk' conception of safety, while understanding the quantification it involves as restricted to eventualities that occur in possible cases close to the actual case. Since safety varies over time, closeness should do likewise. To a first approximation, one is safe in a possible world *w* at a time *t* from an eventuality if and only if that eventuality obtains in no world 'close' to *w* at *t*. Call this the 'no close risk' conception of safety.[9]

Given that every world is always close to itself, the 'no close risk' conception of safety validates arguments A_{safety}. If S was shot in a world *w*, then S was shot in a world close to *w* at any given time *t*, namely *w* itself, and so is not safe in *w* at *t* from being shot. Without need

9 See Sainsbury 1997, Peacocke 1999: 310–28 and Williamson 1992, 1994: 226–30 and 2000: 123–30 for such ideas.

of any such special assumptions about closeness, the 'no close risk' conception of safety also validates argument B_{safety}. For if the premises hold in *w* with reference to a time *t*, then S was not shot by X in any world close to *w* at *t*, S was not shot by Y in any world close to *w* at *t*, S was not shot by Z in any world close to *w* at *t*, and S was not shot other than by X, Y or Z in any world close to *w* at *t*, so S was not shot in any world close to *w* at *t*; thus the conclusion holds in *w* with reference to *t*.

On the 'no close risk' conception, safety is a sort of local necessity, and closeness a sort of accessibility relation between worlds in a possible worlds semantics for modal logic. A_{safety} generalizes to the T axiom schema $\mathbf{A} \rightarrow \Diamond\mathbf{A}$ of modal logic (whatever is is possible), which corresponds to the reflexivity of the accessibility relation. B_{safety} generalizes to the K principle that if $(\mathbf{A}_1 \ \& \ \ldots \ \& \ \mathbf{A}_n) \rightarrow \mathbf{B}$ is valid, so too is $(\Box\mathbf{A}_1 \ \& \ \ldots \ \& \ \Box\mathbf{A}_n) \rightarrow \Box\mathbf{B}$ ($n \geq 0$), which holds for any accessibility relation. Together, the two principles axiomatize the modal system KT (also known as T).

Within this framework, the substantive task remains of specifying closeness as informatively as we can in terms of appropriate respects of similarity, perhaps in context-dependent ways, just as with Lewis's possible worlds semantics for counterfactual conditionals. Later I will discuss closeness for the special case of epistemological safety. For safety in general, I will confine myself to a few brief remarks about closeness and chance.

We should not take for granted that all worlds with a non-zero physical chance in a world *w* at a time *t* of obtaining count as close to *w* at *t*. Of course, that condition holds vacuously if individual worlds are so specific that each of them always has zero chance of obtaining, but then the point must be put in terms of less specific possibilities, where a possibility is a set of worlds and obtains if and only if one of its members does. We should not take for granted that all possibilities with a non-zero chance in a world *w* at a time *t* of obtaining contain worlds that count as close to *w* at *t*. Perhaps, for example, some possibilities may involve such large deviations from *w* in overall trends (short of strict physical laws) that no world in them counts as close to *w* at *t*. Not all worlds with the same laws of nature as *w* that branch from *w* only after *t* need count as close to *w* at *t*.

Recall the case of deterministic worlds, whether or not the actual world is one of them. Let *w* be such a world. If whatever does not happen in *w* was always safe in *w* from happening, then the distinction between safety and danger has no useful application to *w*. In *w*, if you play Russian roulette and get away with it, you were always perfectly safe. That is not the distinction the inhabitants of *w* need for practical purposes. They need one on which playing Russian roulette counts as unsafe, whether or not you get away with it. Presumably, the idea will be something like this: if you play Russian roulette in *w* at *t* and get away with it, you are unsafe in *w* at *t* because in some world *w** relevantly but not perfectly similar to *w*, you play Russian roulette at *t* and do not get away with it. The standard for sufficient similarity must be set reasonably high, otherwise almost everything will count as unsafe, and again the distinction will not be the one the inhabitants of *w* need for practical purposes.

Such a distinction between safety and danger, based on similarity rather than chance, is available in indeterministic worlds too. Even when objective chance is present, it does not automatically 'capture' the distinction. But that does not mean that chance is irrelevant to safety either. For chance can help constitute similarity, both directly through similarity in chances and indirectly through probabilistic laws in virtue of which some further respects of similarity carry more weight than others. But the resultant distinction between safety and danger will not be a directly probabilistic one, because it will have the structural features characteristic of the 'no close risk' conception of safety. In particular, it will validate the arguments A_{safety} and B_{safety}.

In practice, we may expect this non-probabilistic conception of safety to diverge dramatically from a probabilistic conception in at least a few cases, by arguments of a similar form to B_{safety}. Although one can construct formal models in which safety entails a high chance without entailing chance 1 (Williamson 2005: 485–6), it is unclear that all the relevant cases can be modelled in that way. For instance, if we bracket the epistemological aspect of the example of the *n* carpets considered earlier, we may simply ask whether we are safe from the *i*th carpet's not remaining on the floor. By hypothesis, nothing untoward will in fact happen; all the carpets will remain on the floor. For any given *i* a positive answer is plausible; moreover, it would be implausibly *ad hoc* to give a positive answer for some numbers *i* and a negative answer for others. But if for all *i* ($1 \leq i \leq n$) we are safe from the eventuality that the *i*th carpet does not remain on the floor, then by an argument of a similar form to Bsafety we are safe from the eventuality that not every carpet remains on the floor. The latter eventuality will not in fact obtain, but for large enough *n* it has at the relevant time a very high chance of obtaining.

Although this is hardly a comfortable result, the defender of the 'no close risk' conception of safety may be able to live with it. For if we suppose that we are not safe from the eventuality that not every carpet remains on the floor, we cannot consistently suppose in addition that for each carpet we are safe from the eventuality that it does not remain on the floor; the B_{safety}-style argument seems valid even when we deny its conclusion. Thus we infer that for at least one carpet we are not safe from the eventuality that it does not remain on the floor. But since the carpets are all on a par, we conclude that for no carpet are we safe from the eventuality that it does not remain on the floor. Thus denying that we are safe from the eventuality that not every carpet remains on the floor would push us towards the analogue of scepticism for safety, a view on which almost nothing is safe and the distinction between safety and danger becomes useless in practice.

At this point, the probabilist may be inclined to comment: if that is how the 'no close risk' conception of safety works, we are better off without it. Why not simply take advantage of modern technology, and use a probabilistic conception of safety instead?

Undoubtedly, when we have a well-supported probabilistic model of our situation, it is often more prudent to use it than to rely on a 'no close risk' conception of safety. Lotteries and coin-tossing are obvious examples. Of course, a probabilistic model will help us little if it is computationally intractable. In practice, even for lotteries and coin-tossing, we are hardly ever in a position to calculate the quantum mechanical chances, even to within a reasonable approximation. The standard probabilistic models of lotteries and coin-tossing were developed long before quantum mechanics. Their utility does not depend on whether our world is indeterministic. Not only do the models ignore eventualities such as the coin landing on its edge or the lottery being rigged; they need not even approximate the objective chances on a specific occasion. A coin that comes up heads about half the time may nevertheless have an objective chance of 10% or 90% of coming up heads when I toss it right now.

In practice, we must take most of our decisions without the help of a well-supported probabilistic model of our situation.[10] Am I safe from missing my train if I walk to the station?

10 Even if we have well-defined credences (degrees of belief) in the relevant propositions and they satisfy the probability axioms, that is not what is meant by having a well-supported probabilistic model. As already explained, credences are too subjective for purposes of the distinction between safety and danger.

In answering the question, I do not attempt to calculate with probabilities, let alone objective chances. If I did, I could only guess wildly at their values. Moreover, in order to make the calculations, one needs numerical values for the levels of interdependence between the different risks, and I could only guess wildly at their values too. If the result of the calculation was at odds with my non-probabilistic judgment, I might very well adjust my estimates rather than my non-probabilistic judgment.[11]

We need a conception of safety that we can apply quickly in practice, on the basis of vague and impoverished evidence, without making probabilistic calculations.[12] The 'no close risk' conception of safety meets that need. Since it validates B_{safety}-style arguments, it permits us to make ourselves safe from a disjunction of dangers by making ourselves safe from each disjunct separately, and to check that we are safe from the disjunction by checking that we are safe from each disjunct in turn. That way of thinking assumes a closure principle for safety that the 'no close risk' conception can deliver and the 'small risk' conception cannot. The chance of the disjunction is much higher than the chance of any disjunct, but if each disjunct is avoided in all close cases, so is their disjunction. The price of such a practically tractable conception of safety may be that we count as safe from some dangers that have a high chance of obtaining.

Someone might object that safety in this sense cannot be relied on, for when one is safe from being shot even though one has a high chance of being shot, one has a high chance of being simultaneously shot and safe from being shot. But that is a fallacy. The 'no close risk' conception of safety validates argument A_{safety}. If one is safe from being shot, one is not shot; every world is close to itself. Necessarily, if one had been shot, one would not have been safe from being shot. Even though one has a high chance of not being safe from being shot, one is in fact safe from being shot. High chance events do not always occur; that includes safety events. Of course, we sometimes think that we are safe when in fact we are not, but we have no reason to expect ourselves to be infallible about safety, or anything else.

A different objection to the 'no close risk' conception of safety is that it makes safety ungradable, whereas in fact it is gradable – it comes in degrees. It does so even when what

11 I have anecdotal evidence that this happens when the safety of nuclear power stations is estimated.

12 Compare Gigerenzer et al. 1999.

is at issue is someone's being safe from a specific danger at a specific time, the proper analogue of someone's knowing a specific truth at a specific time. For example, we can sensibly ask how safe someone is from being shot, or say that he is safer from being shot by X than from being shot by Y. However, the mere fact that the 'no close risk' conception of safety avoids reliance on a probability threshold does not entail that it makes safety ungradable. It treats safety as a local modality, a restricted sort of necessity. Possibility and necessity do not involve a probability threshold, but they are in some sense gradable. For example, we can sensibly ask how possible or necessary it is to keep children in sight at all times, or say that it is more possible or necessary to do so with children than with cats.[13] There are several ways in which the grading might work. It might concern the proportion of close worlds in which the eventuality obtains, just as a glass two-thirds full is fuller than a glass one-third full even though neither is full; call that the 'proportion view of graded safety'. Alternatively, it might concern the distance from the actual world of the closest worlds in which the eventuality obtains; call that the 'distance view of graded safety'.

Here is an example to illustrate the difference between the two views of graded safety. Your opponent is throwing a die. All six outcomes obtain in equal proportions of close worlds. In the actual world, she will throw a five. On the proportion view of graded safety, you are no safer from her throwing a six than you are from her throwing a five, since the proportion is ⅙ in both cases. But you are safer from her throwing a five than you are from her throwing a number divisible by three, since the proportion is ⅙ in the former case and 2⁄6 in the latter. By contrast, on the distance view of graded safety, you are safer from her throwing a six than you are from her throwing a five, since she throws a five in the actual world, and no counterfactual world is as close to the actual world as the actual world is to itself. You are not safer from her throwing a five than you are from her throwing a number divisible by three, for the same reason.

I will not attempt to decide between the two views of graded safety here. Each is compatible with the 'no close risk' conception of safety. Each has its advantages and disadvantages. They make grading carry different sorts of information. Perhaps we use both, in different

13 Googling the strings "more possible than", "more necessary than", "how possible is" and "how necessary is" yields tens of thousands of examples in each case.

situations. For present purposes the moral is just that grading safety does not undermine the 'no close risk' conception.

The conception of probability as obeying the mathematical principles of the probability calculus goes back only to the mid-seventeenth century.[14] The distinction between safety and danger is far older. No wonder it works according to different principles. But it is no mere survival from pre-modern thinking. It has a distinctive structure of its own that fits it to serve practical purposes for which a probabilistic approach is infeasible. We need both.

6. IT IS TIME TO RETURN to epistemology, and apply the 'no close risk' conception of safety in general to knowledge in particular.

As with safety in general, we may expect the difference in structure between knowledge and high chance to produce dramatic divergences between them in cases specially designed for that effect, such as those constructed by Hawthorne and Lasonen-Aarnio. The alarm that such cases can induce may be lessened by reflection on other examples in which knowledge and high chance come apart. A lottery is about to be drawn. Each ticket has chance $1/n$. Let Lucky be the ticket that will in fact win ('Lucky' is a name, a rigid designator). Lucky has the same chance of winning as any other ticket, namely $1/n$. But we know in advance that Lucky will win.[15] Of course, the tricky linguistic features of such examples make room for manoeuvres that are not available in more straightforward cases, such as the conjunction about the n carpets. Nevertheless, the example shows that any connection between knowledge and high chance would have to be established by very careful argument, not taken as obvious. The present hypothesis is that high chance is not a necessary condition on knowledge; HC fails.

One point of disanalogy between knowledge and safety emerged in the previous section: safety is gradable; knowledge is not. Although we have little difficulty in thinking of some

---

14 Hacking 1975 is the classic work on the emergence of such a mathematical conception of probability. Although subsequent scholarship has modified his account in various ways, they do not concern us here. When I spoke on this material in California, I got an unintended laugh by referring to the mid-seventeenth century as 'recent'.

15 Such examples are of course applications of the account of the contingent *a priori* in Kripke 1980 to objective chance in place of metaphysical necessity. For such applications see Williamson 2006 and Hawthorne and Lasonen-Aarnio 2009.

knowledge as 'more solid' or 'less shaky' than other knowledge, we do not find it natural to express such comparisons by modifying the word 'know' with the usual linguistic apparatus of gradability. This might be a serious objection to a semantic analysis of 'knowledge' in terms of 'safety'. But that is not the project. Rather, the aim is to use safety, in the ordinary sense of 'safety', as a model to help explain the underlying nature of knowledge itself, in the ordinary sense of 'knowledge'.

Clearly, A_{safety}-style arguments correspond to the factiveness of knowledge: if something is so, nobody knows that it is not so. Similarly, B_{safety}-style arguments look as though they should correspond to some sort of multi-premise closure principle.

However, if knowing p is simply being safe from error in the sense of being safe from falsely believing p, the 'no close risk' conception of safety does not automatically predict a multi-premise closure principle such as MPC. For example, suppose that I know p, I know $p \rightarrow q$, and believe q by competent deduction, using *modus ponens*, from those premises. Thus I am safe from falsely believing p, so p is true in all close worlds in which I believe p, and I am safe from falsely believing $p \rightarrow q$, so $p \rightarrow q$ is true in all close worlds in which I believe $p \rightarrow q$. Without extra assumptions, it does not follow that q is true in all close worlds in which I believe q. For, although *modus ponens* preserves truth, there may be close worlds in which I falsely believe q on utterly different grounds, without believing p or $p \rightarrow q$. Thus I do not count as safe from falsely believing q, and so do not count as knowing q.

Such examples are not genuine counter-examples to MPC. I can know q when I believe q on good grounds, even though I might easily have falsely believed q on different grounds. I may know that the Prime Minister was in Oxford today, because I happened to see him, even though he might easily have cancelled, in which case I would still have had the belief, on the basis of the newspaper announcement which I read this morning.

One way to handle such cases is by explicit relativization to bases. For example, suppose that I am safe from falsely believing p on basis b, and safe from falsely believing $p \rightarrow q$ on basis b^*. Thus p is true in all close worlds in which I believe p on basis b, and $p \rightarrow q$ is true in all close worlds in which I believe $p \rightarrow q$ on basis b^*. Let b^{**} be the basis for believing q which consists of believing p on basis b, believing $p \rightarrow q$ on basis b^*, and believing q by competent deduction from those premises. Then in any close world in which I believe q on basis b^{**}, I believe p on basis b and $p \rightarrow q$ on basis b^*, so p and $p \rightarrow q$ are true, so q is true. Thus I am safe from falsely believing q on basis b^{**}.

However, talk of the 'basis' of a belief is much less clear when applied to non-inferential beliefs. In effect, fixing the 'basis' of a belief seems to boil down to requiring some given level of similarity to the actual world in respects relevant to that belief. This suggests that we may be able to reach a tidier and more perspicuous treatment by rolling all such similarities into the overall relation of closeness between worlds. This relation will vary both with the agent and with the time. For simplicity, those two dimensions will be kept fixed and tacit in the formal treatment to come. The idea is that when one knows p 'on basis b', worlds in which one does not believe p 'on basis b' do not count as close; but knowing 'on basis b' requires p to be true in all close worlds in which one believes p 'on basis b'; thus p is true in all close worlds. In this sense, the danger from which one is safe is p's being false, not only one's believing p when it is false.

The simplest formal realization of this idea is a standard possible world semantics for epistemic logic (Hintikka 1962). The formal language is just that of the propositional calculus, augmented with a single unary sentential operator **K**, read 'one knows that'. A model is a triple $<W, R, V>$ where W is a set, whose members we informally conceive as worlds, R is a binary relation (a set of ordered pairs) between members of W, which we informally conceive as the closeness relation between worlds, and V is a function from worlds to sets of atomic formulas, which we informally conceive as the atomic formulas true at the given world. All the epistemology in the model is packed into the relation R; informally, $<w, w^*> \in R$ if and only if w^* is epistemically possible in w, in the sense that whatever the agent knows in w is true in w^*. In the present setting, epistemic possibility is conceived as a form of closeness. Relative to a model of this form, the truth of a formula $\mathbf{A}$ at a world $w \in W$ ($w \vDash \mathbf{A}$) is given a recursive definition of the usual sort; it is displayed here for purposes of later comparison:

$w \vDash \mathbf{p}$	if and only if	$\mathbf{p} \in V(w)$, for atomic $\mathbf{p}$
$w \vDash \mathbf{A} \,\&\, \mathbf{B}$	if and only if	$w \vDash \mathbf{A}$ and $w \vDash \mathbf{B}$
$w \vDash \neg\mathbf{A}$	if and only if	$w \nvDash \mathbf{A}$
$w \vDash \mathbf{KA}$	if and only if	for all $<w, w^*> \in R$, $w^* \vDash \mathbf{A}$

Informally, one knows something if and only if it is true in all close worlds. Formulas $\mathbf{A}_1, \ldots, \mathbf{A}_n$ are said to *entail* a formula $\mathbf{C}$ if and only if with respect to every model $<W, R, V>$ and $w \in W$, if $w \vDash \mathbf{A}_1, \ldots, w \vDash \mathbf{A}_n$ then $w \vDash \mathbf{C}$ (for $n = 0$ this just requires that $w \vDash \mathbf{C}$). Thus entailment is truth-preservation at all worlds in all models.

Since R is interpreted as closeness, a reflexive relation, we are only interested in models in which R is reflexive. As usual, this secures the factiveness of knowledge: **KA** always entails **A**.

Although the semantic clause for **K** does not include a separate belief condition, this does not imply that the intended interpretation permits knowledge without belief. Rather, that interpretation can wrap belief up into the epistemic relation R, so that if the agent does not believe **A** at w, w will have R to some world w^* at which **A** is false, so the agent will not count as knowing **A** at w either. The same goes for other putatively necessary conditions on knowledge.

The notorious problem for the standard possible worlds semantics for epistemic logic is that it validates multi-premise closure in far too strong a form. From the structure of the models, independently of any constraints on R, we have this form of logical omniscience (for $n \geq 0$):

LC If $\mathbf{A}_1, \ldots, \mathbf{A}_n$ entail $\mathbf{C}$ then $\mathbf{KA}_1, \ldots, \mathbf{KA}_n$ entail $\mathbf{KC}$.

This corresponds to the closure condition on safety under the 'no close risks conception'. Thus if one knows some simple truths, one knows any logical consequence of them, no matter how complex and hard to derive. In particular, one automatically knows any logical truth. Similarly, on the corresponding semantics for doxastic logic, if one believes some things, one believes any logical consequence of them, no matter how complex and hard to derive. In particular, one automatically believes any logical truth. Unlike MPC, LC imposes no restriction to cases in which one has carried out the deduction, or even to those in which one has contemplated its conclusion. Out of the frying pan into the fire!

Possible worlds semantics for knowledge attributions has found some diehard defenders, usually amongst those who want to treat the objects of propositional attitudes as sets of possible worlds.[16] As already indicated, simple probabilistic accounts of knowledge have similar problematic consequences for *single*-premise closure, since if *p* entails *q*, the probability of *q* is at least as high as the probability of *p*, even if no one has performed the deduction. When we are studying how deduction *extends* knowledge, such accounts are far too indiscriminate. In particular, if – as here – we are treating knowledge simply as a relation to an object, then we must individuate the objects of knowledge finely enough to permit them to be distinct even when logically equivalent, so that an agent can know *p* without knowing *q*, even though *q* is logically equivalent to *p*. For these purposes, we cannot regard the objects of knowledge as simply sets of possible worlds.

7. A POSSIBLE SOLUTION to the problem of logical omniscience is suggested by a related problem for a safety conception of knowledge. Suppose that S believes a complex truth *p* of first-order logic only on the say-so of his utterly unreliable guru. Surely S does not know *p*, even though *p* itself is perfectly safe from falsity.

The natural response is to note that although S could not have falsely believed *p*, S would just as easily have believed some false counterpart *p** which his guru might easily have intoned instead of *p*. The suggestion is that knowing *p* requires safety from the falsity of *p* and of its epistemic counterparts. The counterpart *p** is close to *p* in a way analogous to that in which a world *w** may be close to a world *w*.[17]

---

16 For a sophisticated and qualified defence of such a view of knowledge and belief attributions see Stalnaker 1984: 71–99 and 1999: 241–73. It is not always appreciated that the account in Lewis 1996 is of this type; indeed, since Lewis's accessibility relation is an equivalence relation, his account validates not only logical omniscience but the very strong epistemic logic S5, with the theorems $\mathbf{KA} \rightarrow \mathbf{KKA}$ and $\neg\mathbf{KA} \rightarrow \mathbf{K}\neg\mathbf{KA}$. The idea that in deduction we gain only the metalinguistic knowledge that a sentence is true is both hopelessly implausible and ad hoc, for example as applied to scientific knowledge, and in any case does not work as a defence of logical omniscience, since the agent typically knows in advance of making the inference elementary semantic facts which, combined with the rest of the agent's knowledge, already entail the truth of the sentence; thus, given logical omniscience, the agent knew in advance of making the inference even that the sentence was true.

17 For the use of epistemic counterparts at the propositional level in safety-style epistemology see Williamson 1994: 231–4 and 2000: 101. Hintikka also proposed using counterparts in the semantics of quantified epistemic logic, although for individuals rather than formulas and in a quite different way from that envisaged here (Hintikka 1970).

We must coordinate counterparthood between different formulas. For consider a different agent who is credulous and undiscriminating with atomic formulas but scrupulously respects logical relations. Suppose that her counterparts for atomic q in worlds w^* and w^{**} are atomic q^* and q^{**} respectively; then her counterparts for $\neg q$ in w^* and w^{**} are $\neg q^*$ and $\neg q^{**}$ respectively. She would never treat $\neg q^{**}$ as the negation of q^* in w^* or $\neg q^*$ as the negation of q^{**} in w^{**}.

We can handle these complexities by replacing the two-place relation R in the models by a three-place relation (for which 'R' will also be used), consisting of triples $<w, w^*, \mathbf{f}>$, where w and w^* are worlds as before but $\mathbf{f}$ is a function mapping all formulas of the language to formulas of the language. The idea is that $\mathbf{f(A)}$ in w^* is a counterpart of $\mathbf{A}$ in w, and that w^* is close to w under the counterpart mapping $\mathbf{f}$. In modelling the example just given, one imposes the special constraint that $\mathbf{f(\neg A)} = \mathbf{\neg f(A)}$ for all formulas $\mathbf{A}$ and triples $<w, w^*, \mathbf{f}>$ in R.

The semantic clause for $\mathbf{KA}$ must be modified in accordance with the use of epistemic counterparts. The new version is:

$$w \models \mathbf{KA} \quad \text{if and only if} \quad \text{for all } <w, w^*, \mathbf{f}> \in R,\ w^* \models \mathbf{f(A)}$$

No other changes in the models are needed, at least for present purposes. Call the new models *refined*.

Here is a toy example to show how knowledge can fail to be closed under even very elementary inferences in refined models. Let W contain a single world w, and R consist of all triples $<w, w, \mathbf{f}>$ such that $\mathbf{f(p \,\&\, p)} = \mathbf{p \,\&\, p}$, where $\mathbf{p}$ is a fixed atomic formula and $\mathbf{p} \in V(w)$, so $w \models \mathbf{p}$ and $w \models \mathbf{p \,\&\, p}$. By the constraint on $\mathbf{f}$, $w \models \mathbf{K(p \,\&\, p)}$. But there is a function $\mathbf{f}$ such that $\mathbf{f(p)} = \mathbf{\neg p}$ and $<w, w, \mathbf{f}> \in R$, so $w \not\models \mathbf{f(p)}$, so, by the semantic clause for $\mathbf{K}$, $w \not\models \mathbf{Kp}$. Thus the agent's knowledge in the model is not even closed under the elementary deduction from $\mathbf{p \,\&\, p}$ to $\mathbf{p}$. Indeed, one can easily check that $\mathbf{p \,\&\, p}$ is *all* that the agent knows in the model.

For unrefined models, we required R to be reflexive – to contain $<w, w>$ for every $w \in W$ – so that $\mathbf{KA}$ would entail $\mathbf{A}$ (the factiveness of knowledge). To achieve the same effect in refined models, we require R to contain $<w, w, \mathbf{1}>$ for every $w \in W$, where $\mathbf{1}$ is the identity function such that $\mathbf{1(A)} = \mathbf{A}$ for every formula $\mathbf{A}$: every formula is a counterpart of itself in the same world. Call refined models meeting this constraint *reflexive*. In such models, if $w \vDash \mathbf{KA}$, then $w \vDash \mathbf{1}(\mathbf{A})$, so $w \vDash \mathbf{A}$.

If we wanted to insist that a formula must be its sole counterpart with respect to the original world, we could add the constraint that $<w, w, \mathbf{f}> \in R$ only if $\mathbf{f} = \mathbf{1}$. Toy models with only one world like that above would then have to be replaced by models with several worlds. Although that could easily be done, for simplicity we may continue to allow R to contain triples $<w, w, \mathbf{f}>$ in which $\mathbf{f} \neq \mathbf{1}$.

Some more general results illustrate the complete failure of logical omniscience in refined models. The logic of reflexive refined models has a sound and complete axiomatization in which the only axioms are truth-functional tautologies and formulas of the form $\mathbf{KA} \rightarrow \mathbf{A}$ and *modus ponens* is the only rule of inference: the models impose no special features on knowledge beyond factiveness. No disjunction of disjuncts of the form $\mathbf{KA}$ is valid in all reflexive refined models: there is no set of formulas at least one of which must be known. Nor does knowing some formulas entail knowing any other formulas in reflexive refined formulas, provided that the formulas do not contain $\mathbf{K}$ (of course factiveness entails theorems such as $\mathbf{KKA} \rightarrow \mathbf{KA}$). All these results are proved in the appendix.

If one wants to guarantee some elementary forms of closure, one can easily do so by adding further constraints on refined models, but they are not intrinsic to the very structure of the semantics.

If we wanted to, we could even recover logical omniscience in a special class of refined models. Given an unrefined model $<W, R, V>$, we can define a corresponding refined model $<W, R\#, V>$ that 'behaves in the same way', where the members of $R\#$ are just the triples $<w, w^*, \mathbf{1}>$ such that $<w, w^*> \in \mathrm{R}$. One can easily check by induction on the complexity of a formula $\mathbf{A}$ that $\mathbf{A}$ has the same truth-value at any world in $<W, R\#, V>$ as it has at that world in $<W, R, V>$. Thus refined models in effect include the original unrefined models as a special case. Consequently, whatever is valid on all refined models is also valid on all unrefined models (but not *vice versa*). Only the inclusion of functions other than $\mathbf{1}$ makes refined models behave differently from refined ones. Given any refined model $<W, S, V>$ such that for all worlds w, w^*, $<w, w^*, \mathbf{f}> \in S$ only if $\mathbf{f} = \mathbf{1}$, we can define a corresponding unrefined model $<W, S\hat{}, V>$, where $<w, w^*> \in S\hat{}$ if and only if $<w, w^*, \mathbf{1}> \in S$. Then any formula has the same truth-value at any world in $<W, S^\wedge, V>$ as it has at that world in $<W, S, V>$. In fact, $S^\wedge\# = S$ and $R\#^\wedge = R$. Call refined models of the form $<W, R\#, V>$ *effectively unrefined.*

Refined models permit the definition of an analogue of competent deduction. The analogue is not a diachronic relation, because far more complex models would be needed to capture changes in safety over time. Rather, it captures a form of synchronic epistemic dependence of a conclusion on premises. The analogue is called *safe derivation*. Given a refined model $<W, R, V>$ and a world $w \in R$, the definition is this:

> $\mathbf{C}$ safely derives at w from $\mathbf{A}_1, \ldots, \mathbf{A}_n$ if and only if whenever $<w, w^*, \mathbf{f}> \in R$, if $w^* \vDash \mathbf{f}(\mathbf{A}_1), \ldots, w^* \vDash \mathbf{f}(\mathbf{A}_n)$ then $w^* \vDash \mathbf{f}(\mathbf{C})$.

We must first check some useful properties of safe derivation. The first is that it preserves truth at any given world in any given reflexive refined model:

> FACT 1 In refined reflexive models:
> If $\mathbf{C}$ safely derives at w from $\mathbf{A}_1, \ldots, \mathbf{A}_n$ and $w \vDash \mathbf{A}_1, \ldots, w \vDash \mathbf{A}_n$ then $w \vDash \mathbf{C}$.

The reason is simply that since the model is reflexive by hypothesis, $<w, w, \mathbf{1}> \in R$, so if $w \vDash \mathbf{1}(\mathbf{A}_1), \ldots, w \vDash \mathbf{1}(\mathbf{A}_n)$ then $w \vDash \mathbf{1}(\mathbf{C})$, in other words, if $w \vDash \mathbf{A}_1, \ldots, w \vDash \mathbf{A}_n$ then $w \vDash \mathbf{C}$. More interestingly, safe derivation preserves *knowledge* at any given world in any given refined model, reflexive or not:

> FACT 2 In refined models:
> If $\mathbf{C}$ safely derives at w from $\mathbf{A}_1, \ldots, \mathbf{A}_n$ and $w \vDash \mathbf{KA}_1, \ldots, w \vDash \mathbf{KA}_n$ then $w \vDash \mathbf{KC}$.

For if $<w, w^*, \mathbf{f}> \in R$ then $w^* \vDash \mathbf{f}(\mathbf{A}_1), \ldots, w^* \vDash \mathbf{f}(\mathbf{A}_n)$ because $w \vDash \mathbf{KA}_1, \ldots, w \vDash \mathbf{KA}_n$, so $w^* \vDash \mathbf{f}(\mathbf{C})$ because $\mathbf{C}$ safely derives at w from $\mathbf{A}_1, \ldots, \mathbf{A}_n$; thus $w \vDash \mathbf{KC}$. Fact 2 is the analogue of MPC for refined models: knowledge is closed under safe derivation.

In models that correspond to the original unrefined ones, safe derivation collapses into knowledge of the material conditional with the conjunction of the premises as antecedent and the conclusion as consequent:

FACT 3 In effectively unrefined models:
C safely derives at w from $\mathbf{A_1}, \ldots, \mathbf{A_n}$ iff $w \vDash \mathbf{K((A_1 \ \& \ \ldots \ \& \ A_n) \rightarrow C)}$.[18]

The proof is trivial. In other refined models that criterion can fail in both directions. Here are two toy examples. As before, for simplicity, W contains a single world, w, and $\mathbf{p} \in V(w)$. First, let $\mathbf{f(p \rightarrow \neg\neg p) = \neg p}$ and $\mathbf{f(A) = A}$ for every formula **A** other than $\mathbf{p \rightarrow \neg\neg p}$, and $R = \{<w, w, 1>, <w, w, \mathbf{f}>\}$. Then at w $\mathbf{\neg\neg p}$ safely derives from **p**, because both formulas are held constant by the functions in R and they are both true at w. But it is not the case that $w \vDash \mathbf{K(p \rightarrow \neg\neg p)}$, since **f** maps that conditional to a formula false at w. Thus the equivalence fails from left to right. Conversely, let $\mathbf{g(\neg\neg p) = \neg p}$ and $\mathbf{g(A) = A}$ for every formula **A** other than $\mathbf{\neg\neg p}$, and $S = \{<w, w, 1>, <w, w, \mathbf{g}>\}$ and consider the model $<W, S, V>$ instead of $<W, R, V>$. Then $w \vDash \mathbf{K(p \rightarrow \neg\neg p)}$, because the functions in S hold $\mathbf{p \rightarrow \neg\neg p}$ constant and it is true at w. But $\mathbf{\neg\neg p}$ does not safely derive at w from **p** in this model, for $w \vDash \mathbf{g(p)}$ but $w \nvDash \mathbf{g(\neg\neg p)}$. Thus the equivalence fails from right to left. Indeed, knowledge is not preserved, for $w \vDash \mathbf{Kp}$ but $w \nvDash \mathbf{K\neg\neg p}$. The same points can be illustrated with more realistic models, but they would be far more complex.

Although safe derivation is not equivalent to knowledge of a conditional, it is a knowledge-like condition, as one can see by comparing its definition with the semantic clause for **KA**. Indeed, **A** is known at a world if and only if it safely derives at that world from the null set of premises. Unlike knowledge of a conditional, safe derivation in refined models is exactly the knowledge-like condition that gets one from knowledge of the premises to knowledge of the conclusion. We might say that safe derivation means that one makes a 'knowledgeable' connection from premises to conclusion, rather than that one knows the connection.

Safe derivation requires an epistemic connection between premises and conclusion, not a logical one. This is already clear in effectively unrefined models. For example, suppose that **p** is known at a world in such a model. Then for any formula **A**, the conditional $\mathbf{(p \leftrightarrow A) \rightarrow A}$ is also known at that world, since it follows from **p**. Thus **A** safely derives from $\mathbf{p \leftrightarrow A}$ at the world, even though **A** is logically independent of $\mathbf{p \leftrightarrow A}$.

18 Define the conjunction of a null set of formulas as a given tautology; then fact 3 holds even when the premise set is null.

Although safe derivation does not require a logical connection between premises and conclusion, it nevertheless possesses the core structural features of a relation of logical consequence – those that hold of all formulas, irrespective of their composition. These features are the Cut Rule (a sort of generalized transitivity), the Rule of Assumptions (reflexivity) and the Thinning Rule (monotonicity):

FACT 4 The Cut Rule holds in refined models:
if **C** safely derives at *w* from $\mathbf{A}_1, \ldots, \mathbf{A}_m$ and **D** safely derives at *w* from $\mathbf{B}_1, \ldots, \mathbf{B}_n$, **C** then **D** safely derives at *w* from $\mathbf{A}_1, \ldots, \mathbf{A}_m, \mathbf{B}_1, \ldots, \mathbf{B}_n$.

FACT 5 The Rule of Assumptions holds in refined models:
A safely derives at *w* from **A**.

FACT 6 The Thinning Rule holds in refined models:
if **C** safely derives at *w* from $\mathbf{A}_1, \ldots, \mathbf{A}_m$ then **C** safely derives at *w* from $\mathbf{A}_1, \ldots, \mathbf{A}_m, \mathbf{B}_1, \ldots, \mathbf{B}_n$.

The proofs of facts 5 and 6 are trivial. To prove fact 4, suppose that **C** safely derives at *w* from $\mathbf{A}_1, \ldots, \mathbf{A}_m$ and **D** safely derives at *w* from $\mathbf{B}_1, \ldots, \mathbf{B}_n$, **C** in some model $<W, R, V>$. Let $<w, w^*, \mathbf{f}> \in R$. If $w^* \models \mathbf{f}(\mathbf{A}_1), \ldots, w^* \models \mathbf{f}(\mathbf{A}_m), w^* \models \mathbf{f}(\mathbf{B}_1), \ldots, w^* \models \mathbf{f}(\mathbf{B}_n)$ then $w^* \models \mathbf{f}(\mathbf{C})$ because **C** safely derives at *w* from $\mathbf{A}_1, \ldots, \mathbf{A}_m$, so $w^* \models \mathbf{f}(\mathbf{D})$ because **D** safely derives at *w* from $\mathbf{B}_1, \ldots, \mathbf{B}_n$, **C**. Thus **D** safely derives at *w* from $\mathbf{A}_1, \ldots, \mathbf{A}_m, \mathbf{B}_1, \ldots, \mathbf{B}_n$, as required. At this level of generality, safe derivation has the same structure as deductive consequence, even though neither implies the other.

To see how constraints on the relation *R* correspond to closure properties of safe derivation, consider the case of conjunction. Call a function **f** *relevant* at a world *w* in a refined model $<W, R, V>$ if and only if $<w, w^*, \mathbf{f}> \in R$ for some $w^* \in W$. When the logical properties of conjunction are 'transparent' to the agent's cognitive system at a world *w*, $\mathbf{f}(\mathbf{A} \,\&\, \mathbf{B}) = \mathbf{f}(\mathbf{A}) \,\&\, \mathbf{f}(\mathbf{B})$ for all formulas **A** and **B** and functions **f** relevant at *w*: the counterpart of a conjunction is the conjunction of the counterparts of its conjuncts. One can easily check that this implies that the introduction and elimination rules for conjunction hold for safe derivation at *w*: **A & B** safely derives at *w* from **A** and **B** and each of **A** and **B** safely derives at *w* from **A & B**. Thus knowledge at *w* is closed under both conjunction introduction and conjunction elimination; one knows a conjunction if and only if one knows the conjuncts.

Of course, it is far more plausible that knowing a conjunction requires knowing the conjuncts than that knowing the conjuncts requires knowing the conjunction; one may fail to put two pieces of knowledge together.[19] However, we should not conclude from this that we need models in which the conjuncts always safely derive from the conjunction, although the conjunction need not safely derive from the conjuncts. Safe derivation is a sufficient but not necessary condition for knowledge to be closed under the given argument. For an agent who has considered the conjuncts separately but not together, and knows the conjuncts but not the conjunction, the conjuncts may not safely derive from the conjunction any more than the conjunction safely derives from the conjuncts. Nevertheless, although the case is a counterexample to the closure of knowledge under conjunction introduction, it is no counterexample to the closure of knowledge under conjunction elimination, precisely because it is not a case of knowing the conjunction. We can model this idea by imposing the weaker constraint that when $w \vDash \mathbf{K(A \,\&\, B)}$, $\mathbf{f(A \,\&\, B) = f(A) \,\&\, f(B)}$: the counterpart of a known conjunction is the conjunction of the counterparts of its conjuncts, but the counterpart of an unknown conjunction need not be. This weaker constraint implies that $w \vDash \mathbf{K(A \,\&\, B)}$ only if $w \vDash \mathbf{KA}$ and $w \vDash \mathbf{KB}$ without implying the converse. It achieves this effect without requiring $\mathbf{A}$ and $\mathbf{B}$ always to safely derive from $\mathbf{A \,\&\, B}$; they do so when the conjunction is known, but may fail to do so otherwise.

A crude, simple way for a model to meet the weaker constraint without meeting the stronger one (that $\mathbf{f(A \,\&\, B) = f(A) \,\&\, f(B)}$ unconditionally) is for it to divide all formulas into two groups, considered (by the agent) and unconsidered (by the agent), relative to each world. If $\mathbf{A \,\&\, B}$ is considered at w, then $\mathbf{f(A \,\&\, B) = f(A) \,\&\, f(B)}$ for every function $\mathbf{f}$ relevant at w. If a formula $\mathbf{C}$ is unconsidered at w, then $\mathbf{f(C)}$ can be anything, in other words, every formula $\mathbf{D}$ is $\mathbf{f(C)}$ function $\mathbf{f}$ relevant at w. Consequently, only considered formulas are known, for if $\mathbf{C}$ is unconsidered at w then for some function $\mathbf{f}$ relevant at w $\mathbf{f(C) = C \,\&\, \neg C}$, so for some world w^* $\langle w, w^*, \mathbf{f} \rangle \in R$ and $w^* \nvDash \mathbf{f(C)}$, so $w \nvDash \mathbf{KC}$. Hence if $w \vDash \mathbf{K(A \,\&\, B)}$ then $\mathbf{A \,\&\, B}$ is considered at w, so $\mathbf{f(A \,\&\, B) = f(A) \,\&\, f(B)}$. Thus such models meet the

19 On the putative closure of knowledge under conjunction elimination see Williamson 2000: 276–83.

weaker constraint, even though they violate the stronger one with respect to unconsidered formulas. So knowing a conjunction implies knowing the conjuncts in models of this type. But knowing the conjuncts does not imply knowing the conjunction in them, for the conjuncts may be separately considered even though their conjunction is unconsidered. Thus we obtain the asymmetry between conjunction introduction and conjunction elimination in a fairly natural way.

None of this poses any problem for the applications of conjunction introduction that were used to raise the problem in section 2. The conjunctions there *are* considered. On the present view, they just are examples that make vivid the radically non-probabilistic nature of knowledge.

These elementary examples suggest that refined models provide a useful safety-inspired framework for analysing the epistemology of inferential relations. No doubt there are subtleties that they cannot capture, but they have enough flexibility for most purposes. Of course, the intended interpretation of the relation *R* has been described only in a highly schematic way. That is hardly surprising, for the whole nature of knowledge is packed into that interpretation. There is still plenty of work to do even at this abstract structural level.

8. A SCIENTIFIC EPISTEMOLOGY needs both a distinction between knowledge and ignorance and a continuum of epistemic probabilities. The latter does not induce a collapse of the former, despite their great structural differences. But it is easier to make structural generalizations about knowledge and about probability than to predict how either of them will apply to individual cases. That is where common sense may be in for a few surprises.

Appendix

Let Σ be the logic axiomatizable with all truth-functional tautologies and all formulas of the form $\mathbf{KA} \rightarrow \mathbf{A}$ as its axioms and *modus ponens* as its rule of inference. We write $\vdash_\Sigma \mathbf{A}$ to mean that $\mathbf{A}$ is a theorem of Σ.

PROPOSITION 1. Σ is sound and complete for the class of reflexive refined models.

Proof: Soundness is obvious (using reflexivity for $\mathbf{KA} \rightarrow \mathbf{A}$). For completeness we use a version of the canonical model construction from modal logic. Let W be the set of all maximal Σ-consistent sets of formulas. Define the epistemic depth of a formula $\mathbf{A}$, ed($\mathbf{A}$), recursively: if $\mathbf{A}$ is atomic, ed($\mathbf{A}$) = 0; ed($\neg\mathbf{A}$) = ed($\mathbf{A}$); ed($\mathbf{A}$ & $\mathbf{B}$) = max{ed($\mathbf{A}$), ed($\mathbf{B}$)}; ed($\mathbf{KA}$) = ed($\mathbf{A}$) + 1. Let R consist of all triples $<w, x, \mathbf{f}>$ such that $w \in W$, $x \in W$, $\mathbf{f}$ is a total function from formulas to formulas, and for every formula $\mathbf{A}$: (1) $\mathbf{KA} \in w$ iff $\mathbf{Kf(A)} \in x$ and (2) ed($\mathbf{A}$) = ed($\mathbf{f(A)}$). R is obviously reflexive. Let $V(w)$ be the intersection of w with the set of atomic formulas. Relative to the reflexive refined model $<W, R, V>$, we prove that for every $w \in W$ and formula $\mathbf{A}$, $w \vDash \mathbf{A}$ iff $\mathbf{A} \in w$ by nested induction on ed($\mathbf{A}$) and within that on the complexity of $\mathbf{A}$. The only interesting case is the induction step for $\mathbf{KA}$. Suppose that for all $w \in W$ and formulas $\mathbf{A}$ such that ed($\mathbf{A}$) $\leq n$, $w \vDash \mathbf{A}$ iff $\mathbf{A} \in w$. Let ed(A) = $\mathbf{n}$. Suppose that $\mathbf{KA} \notin w$. Now $\vdash_\Sigma \neg(\mathbf{A}\ \&\ \neg\mathbf{A})$ and $\vdash_\Sigma \mathbf{K}(\mathbf{A}\ \&\ \neg\mathbf{A}) \rightarrow (\mathbf{A}\ \&\ \neg\mathbf{A})$, so $\vdash_\Sigma \neg\mathbf{K}(\mathbf{A}\ \&\ \neg\mathbf{A})$. Since w is Σ-consistent, $\mathbf{K}(\mathbf{A}\ \&\ \neg\mathbf{A}) \notin w$. Moreover, ed($\mathbf{A}\ \&\ \neg\mathbf{A}$) = ed($\mathbf{A}$). Thus if $\mathbf{f}$ is the function such that $\mathbf{f(A)} = \mathbf{A}\ \&\ \neg\mathbf{A}$ and $\mathbf{f(B)} = \mathbf{B}$ for every other formula $\mathbf{B}$, then $<w, w, \mathbf{f}> \in R$. Since $w \nvDash \mathbf{A}\ \&\ \neg\mathbf{A}$ (= $\mathbf{f(A)}$), $w \nvDash \mathbf{KA}$. For the converse, suppose that $\mathbf{KA} \in w$. Let $<w, w^*, \mathbf{f}> \in R$. Then $\mathbf{Kf(A)} \in w^*$ by definition of R. Since $\vdash_\Sigma \mathbf{Kf(A)} \rightarrow \mathbf{f(A)}$ and w^* is a maximal Σ-consistent set, $\mathbf{f(A)} \in w^*$. But ed($\mathbf{f(A)}$) = ed($\mathbf{A}$) = n by definition of R, so by induction hypothesis, $w^* \vDash \mathbf{f(A)}$. Thus $w \vDash \mathbf{KA}$. This completes the induction. As usual, if $\nvdash_\Sigma \mathbf{A}$ then $\neg\mathbf{A} \in w$ for some $w \in$ W, so $w \nvDash \mathbf{A}$, which gives completeness.

PROPOSITION 2. $\nvdash_{\Sigma}$ $\mathbf{KA_1 \vee \ldots \vee KA}_m$ for any formulas $\mathbf{A_1}, \ldots, \mathbf{A}_m$.
Proof: Consider a model $<W, R, V>$ in which R contains all permissible triples $<w, w^*, \mathbf{f}>$. Pick $w \in W$. For any formula $\mathbf{A}$, $<w, w, \mathbf{f}> \in R$ for some function $\mathbf{f}$ such that $\mathbf{f(A) = A \,\&\, \neg A}$. Since $w \nvDash \mathbf{A \,\&\, \neg A}$, $w \nvDash \mathbf{KA}$. Hence $w \nvDash \mathbf{KA_1 \vee \ldots \vee KA}_m$. Hence $\nvdash_{\Sigma}$ $\mathbf{KA_1 \vee \ldots \vee KA}_m$ by soundness.

PROPOSITION 3. Let $\mathbf{A_1}, \ldots, \mathbf{A}_m, \mathbf{B_1}, \ldots, \mathbf{B_n}$ be $\mathbf{K}$-free formulas. Then $\vdash_{\Sigma}$ $\mathbf{(KA_1 \,\&\, \ldots \,\&\, KA}_m\mathbf{) \rightarrow (KB_1 \vee \ldots \vee KB}_n\mathbf{)}$ iff either $\mathbf{A_1 \,\&\, \ldots \,\&\, A}_m$ is a truth-functional contradiction or for some i, j $\mathbf{A}_i = \mathbf{B}_j$.
Proof: Suppose that $\vdash_{\Sigma}$ $\mathbf{(KA_1 \,\&\, \ldots \,\&\, KA}_m\mathbf{) \rightarrow (KB_1 \vee \ldots \vee KB}_n\mathbf{)}$ but $\mathbf{A_1 \,\&\, \ldots \,\&\, A}_m$ is not a truth-functional contradiction. Then $\mathbf{A_1 \,\&\, \ldots \,\&\, A}_m$ is true under some assignment of truth-values to atomic sentences; let X be the set of atomic sentences true under that interpretation. Construct a model $<W, R, V>$ by setting $W = \{w\}$ for some w, $V(w) = X$, and $<w, w, \mathbf{f}> \in R$ iff for $1 \leq i \leq m$, $\mathbf{f(A}_i\mathbf{)} = \mathbf{A}_i$. By choice of V, $w \vDash \mathbf{A_1 \,\&\, \ldots \,\&\, A}_m$. Thus for $1 \leq i \leq m$, if $<w, w, \mathbf{f}> \in R$ then $w \vDash \mathbf{f(A}_i\mathbf{)}$, so $w \vDash \mathbf{KA}_i$, so $w \vDash \mathbf{KA_1 \,\&\, \ldots \,\&\, KA}_m$. Since the model is reflexive and $\vdash_{\Sigma}$ $\mathbf{(KA_1 \,\&\, \ldots \,\&\, KA}_m\mathbf{) \rightarrow (KB_1 \vee \ldots \vee KB}_n\mathbf{)}$, by soundness $w \vDash \mathbf{KB_1 \vee \ldots \vee KB}_n$, so for some j, $w \vDash \mathbf{KB}_j$. Let $\mathbf{f(B}_j\mathbf{)}$ be a contradiction and $\mathbf{f(A) = A}$ for every other formula $\mathbf{A}$. If $<w, w, \mathbf{f}> \in R$ then $w \vDash \mathbf{f(B}_j\mathbf{)}$, which is impossible. Hence $<w, w, \mathbf{f}> \notin R$, which must be because $\mathbf{B}_j$ is some $\mathbf{A}_i$. Conversely, suppose that $\mathbf{A}_i = \mathbf{B}_j$. Then trivially $\vdash_{\Sigma}$ $\mathbf{(KA_1 \,\&\, \ldots \,\&\, KA}_m\mathbf{) \rightarrow (KB_1 \vee \ldots \vee KB}_n\mathbf{)}$. Finally, if $\mathbf{A_1 \,\&\, \ldots \,\&\, A}_m$ is a truth-functional contradiction then $\vdash_{\Sigma}$ $\mathbf{\neg(A_1 \,\&\, \ldots \,\&\, A}_m\mathbf{)}$; but $\vdash_{\Sigma}$ $\mathbf{(KA_1 \,\&\, \ldots \,\&\, KA}_m\mathbf{) \rightarrow (A_1 \,\&\, \ldots \,\&\, A}_m\mathbf{)}$ since $\vdash_{\Sigma}$ $\mathbf{KA}_i \rightarrow \mathbf{A}_i$ for each i, so $\vdash_{\Sigma}$ $\mathbf{\neg(KA_1 \,\&\, \ldots \,\&\, KA}_m\mathbf{)}$ and therefore $\vdash_{\Sigma}$ $\mathbf{(KA_1 \,\&\, \ldots \,\&\, KA}_m\mathbf{) \rightarrow (KB_1 \vee \ldots \vee KB}_n\mathbf{)}$.

Acknowledgments

This paper is the written version of my Amherst Lecture in Philosophy, delivered at Amherst College on 5 March 2009. Earlier versions of the material were presented to colloquia at Sheffield University, the University of Southern California, Stanford University, a discussion group in Oxford, an Arché workshop at the University of St. Andrews, and the memorial conference for Peter Lipton at the University of Cambridge, where Paul Dicken was the commentator. It develops a theme from my reply (in Williamson 2009) to a paper by John Hawthorne and Maria Lasonen-Aarnio (Hawthorne and Lasonen-Aarnio 2009). It has benefited greatly from discussion on all those occasions. Particular thanks are due to Amherst College for the invitation to deliver the lecture.

References

Carroll, L. 1895. 'What the tortoise said to Achilles'. *Mind* 4: 278–280.

Dretske, F. 1970. 'Epistemic operators'. *Journal of Philosophy* 67: 1007-1023.

Gigerenzer, G., Todd, P.M., and the ABC Research Group (eds.) 1999. *Simple Heuristics That Make Us Smart*. Oxford: Oxford University Press.

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

Hacking, I. 1975. *The Emergence of Probability: A Philosophical Study of Early Ideas about Probability, Induction and Statistical Inference*. Cambridge: Cambridge University Press.

Hájek, A. 2007. 'The reference class problem is your problem too'. *Synthese* 156: 563–585.

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

Hawthorne, J. 2005. 'The case for closure', in Steup and Sosa 2005.

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

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

Hintikka, K.J.J. 1970. 'Objects of knowledge and belief: acquaintances and public figures', *Journal of Philosophy* 67: 869–883.

Kripke, S.A. 1980. *Naming and Necessity*. Oxford: Blackwell.

Kvanvig, J. (ed.) 1996. *Warrant in Contemporary Epistemology*. Lanham, Md.: Rowman & Littlefield.

Kyburg, Jr, H.E. 1961. *Probability and the Logic of Rational Belief*. Middletown: Wesleyan University Press.

Lasonen-Aarnio, M. 2008. 'Single-premise deduction and risk', *Philosophical Studies* 141: 157–173.

Lewis, D.K. 1973. *Counterfactuals*. Oxford: Blackwell.

Lewis, D.K. 1979. 'Counterfactual dependence and time's arrow'. *Noûs* 13: 455–476. Reprinted with postscripts in Lewis 1986.

Lewis, D.K. 1986. *Philosophical Papers, Volume II*. Oxford: Oxford University Press.

Lewis, D.K. 1996. 'Elusive knowledge'. *Australasian Journal of Philosophy* 74: 549–567.

Makinson, D.C. 1965. 'The paradox of the preface', *Analysis* 25: 205–207.

Nozick, R. 1981. *Philosophical Explanations*. Oxford: Oxford University Press.

Peacocke, C.A.B. 1999. *Being Known*. Oxford: Clarendon Press.

Roth, M., and Ross, G. (eds.) 1990. *Doubting*. Dordrecht: Kluwer.

Sainsbury, R.M. 1997. 'Easy possibilities', *Philosophy and Phenomenological Research* 57: 907–919.

Smith, Q. (ed.) 2008. *Epistemology: New Essays*. Oxford: Oxford University Press.

Sosa, E. 1996. 'Postscript to "Proper functionalism and virtue epistemology"', in Kvanvig 1996.

Sosa, E. 2000. 'Contextualism and skepticism'. *Philosophical Issues* 10: 1–18.

Sosa, E. 2007. *A Virtue Epistemology: Apt Belief and Reflective Knowledge*. Oxford: Clarendon Press.

Stalnaker, R.C. 1984. *Inquiry*. Cambridge, Mass.: MIT Press.

Stalnaker, R.C. 1999. *Context and Content*. Oxford University Press.

Stanley, J. 2005. *Knowledge and Practical Interests*. Oxford: Clarendon Press.

Steup, M., and Sosa, E. (eds.) 2005. *Contemporary Debates in Epistemology*. Oxford: Blackwell.

Vogel, J. 1990. 'Are there counter-examples to the closure principle?', in Roth and Ross 1990.

Williamson, T. 1992. 'Inexact knowledge'. *Mind* 101: 217-242.

Williamson, T. 1994. *Vagueness*. London: Routledge.

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

Williamson, T. 2005. 'Replies to commentators'. *Philosophy and Phenomenological Research* 70: 468–491.

Williamson, T. 2006. 'Indicative versus subjunctive conditionals, congruential versus non-hyperintensional contexts'. *Philosophical Issues* 16: 310-333.

Williamson, T. 2007. 'How probable is an infinite sequence of heads?' *Analysis* 67: 173–180.

Williamson, T. 2008. 'Why epistemology cannot be operationalized', in Smith 2008.

Williamson, T. 2009. 'Replies to critics', in Greenough and Pritchard 2009.