To appear in *Grazer Philosophische Studien* 2007 and in Christian Beyer and Alex Burri (eds.), *Philosophical Knowledge — Its Possibility and Scope*, Amsterdam: Rodopi.

Philosophical Knowledge and Knowledge of Counterfactuals[*]

Timothy Williamson

ABSTRACT

Metaphysical modalities are definable from counterfactual conditionals, and the epistemology of the former is a special case of the epistemology of the latter. In particular, the role of conceivability and inconceivability in assessing claims of possibility and impossibility can be explained as a special case of the pervasive role of the imagination in assessing counterfactual conditionals, an account of which is sketched. Thus scepticism about metaphysical modality entails a more far-reaching scepticism about counterfactuals. The account is used to question the significance of the distinction between *a priori* and *a posteriori* knowledge.

§0. Philosophers characteristically ask not just whether things are some way but whether they could have been otherwise. What could have been otherwise is *metaphysically contingent*; what could not is *metaphysically necessary*. We have some knowledge of such matters. We know that Henry VIII could have had more than six wives, but that three plus three could not have been more than six. So there should be an epistemology of metaphysical modality.

The differences between metaphysical necessity, contingency and impossibility are not mind-dependent, in any useful sense of that tantalizing phrase. Thus they are not differences in actual or potential psychological, social, linguistic or even epistemic status (Kripke 1980 makes the crucial distinctions). One shortcut to this conclusion uses the plausible idea that mathematical truth is mind-independent. Since mathematics is not contingent, the difference between truth and falsity in mathematics is also the difference between necessity and impossibility; consequently, the difference between necessity and impossibility is mind-independent. The difference between contingency and non-contingency is equally mind-independent; for if C is a mind-independently true or false mathematical conjecture, then one of C and its negation conjoined with the proposition that Henry VIII had six wives forms a contingently true conjunction while the other forms an impossible conjunction, but which is which is mind-independent. To emphasize the point, think of the mind-independently truth-valued conjecture as evidence-transcendent, absolutely undecidable, neither provable nor refutable by any means. Thus the epistemology of metaphysical modality is one of mind-independent truths.

Nevertheless, doubts begin to arise. Although philosophers attribute metaphysical necessity to mathematical theorems, what matters mathematically is just their truth, not their metaphysical necessity: mathematics does not need the concept of metaphysical necessity. Does metaphysical modality really matter outside philosophy? Even if physicists care about the physical necessity of the laws they conjecture, does it matter to physics whether physically necessary laws are also metaphysically necessary? In ordinary life, we care whether someone could have done otherwise, or whether disaster could have been averted, but the kind of possibility at issue there is far more narrowly circumscribed than metaphysical possibility, by not prescinding from metaphysically contingent initial conditions. He could not have done otherwise because he was in chains, even though it was metaphysically contingent that he was in chains. Does 'could have been' ever express metaphysical possibility when used non-philosophically?

If thought about metaphysical modality is the exclusive preserve of philosophers, so is knowledge of metaphysical modality. The epistemology of metaphysical modality tends to be treated as an isolated case. For instance, much of the discussion concerns how far, if at all, conceivability is a guide to possibility, and inconceivability to impossibility (Gendler and Hawthorne 2002 has a sample of recent contributions to this debate). The impression is that, outside philosophy, the primary cognitive role of conceiving is propaedeutic. Conceiving a hypothesis is getting it onto the table, putting it up for serious consideration as a candidate for truth. The inconceivable never even gets that far. Conceivability is certainly no good evidence for the restricted kinds of possibility that we care about in natural science or ordinary life. We easily conceive particles violating what are in fact physical laws, or the man without his chains. On this view, conceiving, outside philosophy, is not a faculty for distinguishing between truth and falsity in some domain, but rather a preliminary to any such faculty. Although there are truths and falsehoods about conceivability and inconceivability, they concern our mental capacities, whereas metaphysical modalities are supposed to be mind-independent. They are not contingent on mental capacities, because not contingent on anything (at least if we accept the principles of the modal logic S5, that the necessary is necessarily necessary and the possible necessarily possible). When philosophers present conceiving as a faculty for distinguishing between truth and falsity in the domain of metaphysical modality, that looks suspiciously like some sort of illicit projection or unacknowledged fiction: at best, attributions of metaphysical modality would lack the cognitive status traditionally ascribed to them (compare Blackburn 1987; Craig 1985; Wright 1989). The apparent cognitive isolation of metaphysically modal thought makes such suspicions hard to allay. Presenting it as *sui generis* suggests that it can be surgically removed from our conceptual scheme without collateral damage. If it can, what good does it do us? In general, the postulation by philosophers of a special cognitive capacity exclusive to philosophical or quasi-philosophical thinking looks like a scam.

Humans evolved under no pressure to do philosophy. Presumably, survival and reproduction in the stone age depended little on philosophical prowess (dialectical skill was probably no more effective then as a seduction technique than it is now). Any cognitive capacity that we have for philosophy is a more or less accidental byproduct of other developments. Nor are psychological dispositions that are non-cognitive outside philosophy likely suddenly to become cognitive within it. We should expect the cognitive capacities used in philosophy to be cases of general cognitive capacities used in ordinary life, perhaps trained, developed and systematically applied in various special ways, just as the cognitive capacities that we use in mathematics and natural science are rooted in more primitive cognitive capacities to perceive, count, reason, discuss …. In particular, a plausible non-sceptical epistemology of metaphysical modality should subsume our capacity to discriminate metaphysical possibilities from metaphysical impossibilities under more general cognitive capacities used in ordinary life.

I will argue that the ordinary cognitive capacity to handle counterfactual conditionals carries with it the cognitive capacity to handle metaphysical modality. §1 illustrates with examples our cognitive use of counterfactual conditionals. §2 sketches the beginnings of an epistemology of such conditionals. §3 explains how they subsume metaphysical modality. §4 discusses some objections. §5 briefly raises the relation between metaphysical possibility and the restricted kinds of possibility that seem more relevant to ordinary life. Philosophers' ascriptions of metaphysical modality are far more deeply rooted in our ordinary cognitive practices than most sceptics realize.

§1. We start with a well-known example that proves the term 'counterfactual conditional' misleading. As Alan Ross Anderson pointed out (1951: 37), a doctor might say:

(1) If Jones had taken arsenic, he would have shown just exactly those symptoms which he does in fact show.

Clearly, (1) can provide abductive evidence by inference to the best explanation for its antecedent (see Edgington 2003: 23-7 for more discussion):

(2) Jones took arsenic.

If further tests subsequently verify (2), they confirm the doctor's statement rather than in any way falsifying it or making it inappropriate. If we still call subjunctive conditionals like (1) 'counterfactuals', the reason is not that they imply or presuppose the falsity of their antecedents.

Of course, what (2) explains is not the trivial necessary truth that Jones shows whatever symptoms he shows. What is contingent is that Jones shows exactly those symptoms which he does in fact show — he could have shown other symptoms, or none — and, given (1), (2) explains that contingent truth.

While (1) provides valuable empirical evidence, the corresponding indicative conditional does not (Stalnaker 1999: 71):

(1I) If Jones took arsenic, he shows just exactly those symptoms which he does in fact show.

We can safely assent to (1I) without knowing what symptoms Jones shows, since it holds whatever they are. Informally, (1) is non-trivial because it depends on a comparison between independently specified terms, the symptoms which Jones would have shown if he had taken arsenic and the symptoms which he does in fact show; by contrast, (1I) is trivial because it involves only a comparison of his symptoms with themselves. Thus the process of evaluating the 'counterfactual' conditional requires something like two files, one for the actual situation, the other for the counterfactual situation, even if these situations turn out to coincide. No such cross-comparison of files is needed to evaluate the indicative conditional. Of course, when one evaluates an indicative conditional while disbelieving its antecedent, one must not confuse one's file of beliefs with one's file of judgments on the supposition of the antecedent, but that does not mean that cross-referencing from the latter file to the former can play the role that it did in the counterfactual case.

Since (1) constitutes empirical evidence, its truth was not guaranteed in advance. If Jones had looked suitably different, the doctor would have had to assert the opposite counterfactual conditional:

(3) If Jones had taken arsenic, he would not have shown just exactly those symptoms which he does in fact show.

From (3) we can deduce the falsity of its antecedent. For modus ponens is generally agreed to be valid for counterfactual conditionals. Thus (2) and (3) yield:

(4) Jones does not show just exactly those symptoms which he does in fact show.

Since (4) is obviously false, we can deny (2) given (3).

The indicative conditional corresponding to (3) is:

(3I) If Jones took arsenic, he does not show just exactly those symptoms which he does in fact show.

To assert (3I) would be like saying ‘If Jones took arsenic, pigs can fly’. Although a very confident doctor might assert (3I), on the grounds that Jones certainly did not take arsenic, that certainty may in turn be based on confidence in (3), and therefore on the comparison of actual and counterfactual situations.

Could a Bayesian account dispense with the counterfactual conditionals in favour of conditional probabilities? Consider the simple case in which we completely trust the doctor who asserts (1). Before the doctor speaks, we are certain what symptoms Jones shows but agnostic over the characteristic symptoms of arsenic poisoning. We want to update our probability for his having taken arsenic on evidence from the doctor, in Bayesian terms by conditionalizing on it. The doctor cannot simply tell us what probability to assign, because we may have further relevant evidence unavailable to the doctor, for example about Jones’s character. We need the doctor to say something which we can use as evidence; (1) exactly fits the bill (of course, our evidence also includes the fact that the doctor asserted (1), but in the circumstances we can treat (1) itself as the relevant part of our evidence). It may even do better than a non-modal generalization such as ‘Jones showed exactly those symptoms which everyone who takes arsenic shows’: for the symptoms may vary with bodily characteristics of the victim, and through long experience the doctor may be able to judge what symptoms Jones would have shown if he had taken arsenic without being able to articulate a suitable generalization. Any Bayesian account depends on an adequately varied stock of propositions to act as bearers of probability, as evidence or hypotheses. Sometimes that range has to include counterfactual conditionals.

We also use the notional distinction between actual and counterfactual situations to make evaluative comparisons:

(5) If Jones had not taken arsenic, he would have been in better shape than he now is.

Such counterfactual reflections facilitate learning from experience; one may decide never to take arsenic oneself. Formulating counterfactuals about past experience is empirically correlated with improved future performance in various tasks.[1]

Evidently, counterfactual conditionals give clues to causal connections. This point does not commit one to the ambitious programme of analysing causality in terms of counterfactual conditionals (Lewis 1973b, Collins, Hall and Paul 2004), or counterfactual conditionals in terms of causality (Jackson 1977). If the former programme succeeds, all causal thinking is counterfactual thinking; if the latter succeeds, all counterfactual thinking is causal thinking. Either way, the overlap is so large that we cannot have one without much of the other. It may well be over-optimistic to expect either necessary and sufficient conditions for causal statements in counterfactual terms or necessary and sufficient conditions for counterfactual statements in causal terms. Even so, counterfactuals surely play a crucial role in our causal thinking (see Harris 2000: 118-139 and Byrne 2005: 100-128 for some empirical discussion). Only extreme sceptics deny the cognitive value of causal thought.

At a more theoretical level, claims of nomic necessity support counterfactual conditionals. If it is a law that property P implies property Q, then typically if something were to have P, it would have Q. If we can falsify the counterfactual in a specific case, perhaps by using better-established laws, we thereby falsify that claim of lawhood. We sometimes have enough evidence to establish what the result of an experiment would be without actually doing the experiment: that matters in a world of limited resources.

Counterfactual thought is deeply integrated into our empirical thought in general. Although that consideration will not deter the most dogged sceptics about our knowledge of counterfactuals, it indicates the difficulty of preventing such scepticism from generalizing implausibly far, since our beliefs about counterfactuals are so well-integrated into our general knowledge of our environment. I proceed on the assumption that we have non-trivial knowledge of counterfactuals.

§2. In discussing the epistemology of counterfactuals, I assume no particular theory of their compositional semantics, although I sometimes use the Stalnaker-Lewis approach for purposes of illustration and vividness. That evasion of semantic theory might seem dubious, since it is the semantics which determines what has to be known. However, we can go some way on the basis of our pretheoretical understanding of such conditionals in our native language. Moreover, the best developed formal semantic theories of counterfactuals use an apparatus of possible worlds or situations at best distantly related to our actual cognitive processing. While that does not refute such theories, which concern the truth-conditions of counterfactuals, not how subjects attempt to find out whether those truth-conditions obtain, it shows how indirect the relation between the semantics and the epistemology may be. When we come to fine-tune our epistemology of counterfactuals, we may need an articulated semantic theory, but at a first pass we can make do with some sketchy remarks about their epistemology while remaining neutral over their deep semantic analysis. As for the psychological study of the processes underlying our assessment of counterfactual conditionals, it remains in a surprisingly undeveloped state, as recent authors have complained (Evans and Over 2004: 113-131).

Start with an example. You are in the mountains. As the sun melts the ice, rocks embedded in it are loosened and crash down the slope. You notice one rock slide into a bush. You wonder where it would have ended if the bush had not been there. A natural way to answer the question is by visualizing the rock sliding without the bush there, then bouncing down the slope. You thereby come to know this counterfactual:

(6) If the bush had not been there, the rock would have ended in the lake.

You could test that judgment by physically removing the bush and experimenting with similar rocks, but you know (6) even without performing such experiments. Semantically, the counterfactual about the past is independent of claims about future experiments (for a start, the slope is undergoing continual small changes).

Somehow, you came to know the counterfactual by using your imagination. That sounds puzzling if one conceives the imagination as unconstrained. You can imagine the rock rising vertically into the air, or looping the loop, or sticking like a limpet to the slope. What constrains imagining it one way rather than another?

You do not imagine it those other ways because your imaginative exercise is radically informed and disciplined by your perception of the rock and the slope and your sense of how nature works. The default for the imagination may be to proceed as 'realistically' as it can, subject to whatever deviations the thinker imposes by brute force:

here, the absence of the bush. Thus the imagination can in principle exploit all our background knowledge in evaluating counterfactuals. Of course, how to separate background knowledge from what must be imagined away in imagining the antecedent is Goodman's old, deep problem of cotenability (1955). For example, why don't we bring to bear our background knowledge that the rock did not go far, and imagine another obstacle to its fall? Difficult though the problem is, it should not make us lose sight of our considerable knowledge of counterfactuals: our procedures for evaluating them cannot be too wildly misleading.

Can the imaginative exercise be regimented as a piece of reasoning? We can undoubtedly assess some counterfactuals by straightforward reasoning. For instance:

(7) If twelve people had come to the party, more than eleven people would have come to the party.

We can deduce the consequent 'More than eleven people came to the party' from the antecedent 'Twelve people came to the party', and assert (7) on that basis. Similarly, it may be suggested, we can assert (6) on the basis of inferring its consequent 'The rock ended in the lake' from the premise 'The bush was not there', given auxiliary premises about the rock, the mountainside and the laws of nature.

At the level of formal logic, we have the corresponding plausible and widely accepted closure principle that, given a derivation of $\mathbf{C}$ from $\mathbf{B}_1$, …, $\mathbf{B}_{\boldsymbol{n}}$, we can derive the counterfactual conditional $\mathbf{A} \,\Box\!\!\rightarrow \mathbf{C}$ from the counterfactual conditionals $\mathbf{A} \,\Box\!\!\rightarrow \mathbf{B}_1$, …., $\mathbf{A} \,\Box\!\!\rightarrow \mathbf{B}_{\boldsymbol{n}}$; in other words, the counterfactual consequences of a supposition $\mathbf{A}$ are closed under logical consequence (Lewis calls this ‘Deduction within Conditionals’, 1986: 132). With the uncontroversial reflexivity principle **A □→ A**, it follows that, given a derivation of **C** from **A** alone, we can derive **A □→ C** from the null set of premises.

We cannot automatically extend the closure rule to the case of auxiliary premises, for since we can derive an arbitrary conclusion **C** from an arbitrary premise **A** with **C** as auxiliary premise, we could then derive **A □→ C** from the auxiliary premise **C** alone: but that is in effect the invalid principle that any truth is a counterfactual consequence of any supposition whatsoever. Auxiliary premises cannot always be copied into the scope of counterfactual suppositions (the problem of cotenability again).

Even with this caution, the treatment of the process by which we reach counterfactual judgments as inferential is problematic in several ways.

First, a technical problem: not every inference licenses us to assert the corresponding counterfactual, even when the inference is deductive and the auxiliary premises are selected appropriately. For the consequent of (1) is a logical truth (count it vacuously true if Jones shows no symptoms):

(8) Jones shows just exactly those symptoms which he does in fact show.

Thus (8) follows from any premises, including (2), the antecedent of (1); but we cannot assert (1) on the basis of that trivial deduction alone, independently of *which* symptoms Jones does in fact show. This is related to Kaplan’s point that the rule of necessitation fails in languages with terms such as ‘actually’ (1989). The logical truth of (8) does not guarantee the logical truth, or even truth, of (9):

(9) It is necessary that Jones shows just exactly those symptoms which he does in fact show.

For it is contingent that Jones shows just exactly those symptoms which he does in fact show.[2] But let us assume that this technical problem can be solved by a restriction on the type of reasoning from antecedent to consequent that can license a counterfactual, and on the closure principle above, like the restriction on the type of reasoning that licenses the necessitation of its conclusion.

A more serious problem is that the putative reasoner may lack general-purpose cognitive access to the auxiliary premises of the putative reasoning. In particular, the required folk physics may be stored in the form of some analogue mechanism, perhaps embodied in a connectionist network, which the subject cannot articulate in propositional form. Normally, a subject who uses negation and derives a conclusion from some premises can at least entertain the negation of a given premise, whether or not they are willing to assert it, perhaps on the basis of the other premises and the negation of the conclusion. Our reliance on folk physics does not enable us to entertain its negation. This strains the analogy with explicit reasoning.

The third problem is epistemological. Normally, someone who believes a conclusion on the sole basis of deduction from some premises knows the conclusion only if they know the premises. As a universally generalized theory, folk physics is presumably strictly speaking false: its predictions are inaccurate in some circumstances. Consequently, it is not known. But the conclusion that no belief formed on the basis of folk physics constitutes knowledge is wildly sceptical. For folk physics is reliable enough in many circumstances to be used in the acquisition of knowledge, for example that the cricket ball will land in that field. Thus we should not conceive folk physics as a premise of that conclusion. Nor should we conceive some local fragment of folk physics as the premise. For it would be quite unmotivated to take an inferential approach overall while refusing to treat this local fragment as itself derived from the general theory of folk physics. We should conceive folk physics as a locally but not globally reliable method of belief formation, not as a premise.

The preceding reasons motivate the attempt to understand the imaginative exercises by which we judge counterfactuals like (6) as not purely inferential. An attractive suggestion is that some kind of simulation is involved: the difficulty is to explain what that means. It is just a hint of an answer to say that in simulation cognitive faculties are run off-line. The cognitive faculties that would be run on-line to evaluate **A** and **B** as free-standing sentences are run off-line in the evaluation of the counterfactual conditional **A** □→ **B**.[3] This suggests that the cognition has a roughly compositional structure. Our capacity to handle **A** □→ **B** embeds our capacities to handle **A** and **B**, and our capacity to handle the counterfactual conditional operator involves a general capacity to go from capacities to handle the antecedent and the consequent to a capacity to handle the whole conditional. Here the capacity to handle an expression generally comprises more than mere linguistic understanding of it, since it involves ways of assessing its application that are not built into its meaning. But it virtually never involves a decision procedure that enables us always to determine the truth-values of every sentence in which the expression principally occurs, since we lack such decision procedures. Of course, we can sometimes take shortcuts in evaluating counterfactual conditionals. For instance, we can know that $\mathbf{A} \mathbin{\Box\!\!\rightarrow} \mathbf{A}$ is true even if we have no idea how to determine whether $\mathbf{A}$ is true. Nevertheless, the compositional structure just described seems more typical.

How do we advance from capacities to handle the antecedent and the consequent to a capacity to handle the whole conditional? ‘Off-line’ suggests that the most direct links with perception have been cut, but that vague negative point does not take us far. Perceptual input is crucial to the evaluation of counterfactuals such as (1) and (6).

The best developed simulation theories concern our ability to simulate the mental processes of other agents (or ourselves in other circumstances), putting ourselves in their shoes, as if thinking and deciding on the basis of their beliefs and desires (see for example Davies and Stone 1995, Nichols and Stich 2003). Such cognitive processes may well be relevant to the evaluation of counterfactuals about agents. Moreover, they would involve just the sort of constrained use of the imagination indicated above. How would Mary react if you asked to borrow her car? You could imagine her immediately shooting you, or making you her heir; you could even imagine reacting like that from her point of view, by imagining having sufficiently bizarre beliefs and desires. But you do not. Doing so would not help you determine how she really would react. Presumably, what you do is to hold fixed her actual beliefs and desires (as you take them to be just before the request); you can then imagine the request from her point of view, and think through the scenario from there. Just as with the falling rock, the imaginative exercise is richly informed and disciplined by your sense of what she is like.

How could mental simulation help us evaluate a counterfactual such as (6), which does not concern an agent? Even if you somehow put yourself in the rock’s shoes, imagining first-personally being that shape, size and hardness and bouncing down that slope, you would not be simulating the rock's reasoning and decision-making. Thinking of the rock as an agent is no help in determining its counterfactual trajectory. A more natural way to answer the question is by imagining third-personally the rock falling as it would visually appear from your actual present spatial position; you thereby avoid the complex process of adjusting your current visual perspective to the viewpoint of the rock. Is that to simulate the mental states of an observer watching the rock fall from your present position?[4] By itself, that suggestion explains little. For how do we know what to simulate the observer seeing next? But that question is not unanswerable. For we have various propensities to form expectations about what happens next: for example, to project the trajectories of nearby moving bodies into the immediate future (otherwise we could not catch balls). Perhaps we simulate the initial movement of the rock in the absence of the bush, form an expectation as to where it goes next, feed the expected movement back into the simulation as seen by the observer, form a further expectation as to its subsequent movement, feed that back into the simulation, and so on. If our expectations in such matters are approximately correct in a range of ordinary cases, such a process is cognitively worthwhile. The very natural laws and causal tendencies which our expectations roughly track also help to determine which counterfactual conditionals really hold.

However, talk of simulating the mental states of an observer may suggest that the presence of the observer is part of the content of the simulation. That does not fit our evaluation of counterfactuals. Consider:

(10) If there had been a tree on this spot a million years ago, nobody would have known.

Even if we visually imagine a tree on this spot a million years ago, we do not automatically reject (10) because we envisage an observer of the tree. We may imagine the tree as having a certain visual appearance from a certain viewpoint, but that is not to say that we imagine it as appearing to someone at that viewpoint. For example, if we imagine the sun as shining from behind that viewpoint, by imagining the tree's shadow stretching back from the tree, we are not obliged to imagine either the observer's shadow stretching towards the tree or the observer as perfectly transparent.[5] Nor, when we consider (10), are we asking whether if we had believed that there was a tree on this spot a million years ago, we would have believed that nobody knew.[6] It may be better not to think of the simulation as specifically *mental* simulation at all.

Of course, for many counterfactuals the relevant expectations are not hardwired into us in the way that those concerning the trajectories of fast-moving objects around us may need to be. Our knowledge that if a British general election had been called in 1948 the Communists would not have won may depend on an off-line use of our capacity to predict political events. Still, where our more sophisticated capacities to predict the future are reliable, so should be corresponding counterfactual judgments. In these cases too, simulating the mental states of an imaginary observer seems unnecessary.

The off-line use of expectation-forming capacities to judge counterfactuals corresponds to the widespread picture of the semantic evaluation of those conditionals as 'rolling back' history to shortly before the time of the antecedent, modifying its course by stipulating the truth of the antecedent and then rolling history forward again according to patterns of development as close as possible to the normal ones to test the truth of the consequent (compare Lewis 1979). Not all counterfactual conditionals can be so evaluated, since the antecedent need not concern a limited time: in evaluating the claim that space-time has ten dimensions, a scientist can sensibly ask whether if it were true the actually observed phenomena would have occurred. Explicit reasoning may play a much larger role in the evaluation of such conditionals.

Reasoning and prediction do not exhaust our capacity to evaluate counterfactuals. If twelve people had come to the party, would it have been a large party? To answer, one does not imagine a party of twelve people and then predict what would happen next. The question is whether twelve people would have constituted a large party, not whether they would have caused one. Nor is the process of answering best conceived as purely inferential, if one has no special antecedent beliefs as to how many people constitute a large party, any more than the judgment whether the party is large is purely inferential when made at the party. Rather, in both cases one must make a new judgment, even though it is informed by what one already believes or imagines about the party. To call the new judgment 'inferential' simply because it is not made independently of all the thinker's prior beliefs or suppositions is to stretch the term 'inferential' beyond its useful span. At any rate, the judgment cannot be derived from the prior beliefs or suppositions purely by the application of general rules of inference. For example, even if you have the prior belief that a party is large if and only if it is larger than the average size of a party, in order to apply it to the case at hand you also need to have a belief as to what the average size of a party is; if you have no prior belief as to that, and must form one by inference, an implausible regress threatens, for you do not have the statistics of parties in your head. Similarly, if you try to judge whether this party is large by projecting inductively from previous judgments as to whether parties were large, that only pushes the question back to how those previous judgments were made.

In general, our capacity to evaluate counterfactuals recruits *all* our cognitive capacities to evaluate sentences. A quick proof of this uses the assumption that a counterfactual with a true antecedent has the same truth-value as its consequent, for then any sentence $\mathbf{A}$ is logically equivalent to $\mathbf{T} \mathrel{\Box\!\!\rightarrow} \mathbf{A}$, where $\mathbf{T}$ is a trivial tautology; so any serious cognitive work needed to evaluate $\mathbf{A}$ is also needed to evaluate $\mathbf{T} \mathrel{\Box\!\!\rightarrow} \mathbf{A}$.[7]

We can schematize the process of evaluating a counterfactual conditional thus: the thinker imaginatively supposes the antecedent and counterfactually develops the supposition, adding further judgments within the supposition by reasoning, off-line predictive mechanisms and other off-line judgments. To a first approximation: if the development eventually leads us to add the consequent, we assent to the conditional; if not, we dissent from it. Of course, this initial sketch is much too crude, in several ways. We may not be confident enough about the background conditions to decide for or against the conditional. Even if we are confident enough in that respect, if the consequent has not emerged after a given period of development the question remains whether it will emerge in the course of further development, for lines of reasoning can be continued indefinitely from any given premise. To reach a negative conclusion, we must in effect judge that if the consequent were ever going to emerge it would have done so by now (for example, we may have been smoothly fleshing out a scenario incompatible with the consequent with no hint of difficulty). A further over-simplification was that we develop the initial supposition only once: if we find various different ways of imagining the antecedent holding equally good, we may try developing several of them, to see whether they all yield the consequent. For example, if in considering (10) you initially imagine a palm tree, you do not immediately judge that if there had been a tree on this spot a million years ago it would have been a palm tree, because you know that you can equally easily imagine a fir tree. Although far more needs to be said, these remarks may at least start us in the right direction.

Despite its discipline, our imaginative evaluation of counterfactual conditionals is manifestly fallible. We can easily misjudge their truth-values, through background ignorance or error, and distortions of judgment. But such fallibility is the common lot of human cognition. Our use of the imagination in evaluating counterfactuals is practically indispensable. Rather than cave in to scepticism, we should admit that our methods sometimes yield knowledge of counterfactuals.

§3. How does the epistemology of counterfactual conditionals bear on the epistemology of metaphysical modality? We can approach this question by formulating two plausible constraints on the relation between counterfactual conditionals and metaphysical modalities. Henceforth, ‘necessary’ and ‘possible’ will be used for the metaphysical modalities unless otherwise stated.

First, the strict conditional implies the counterfactual conditional:

NECESSITY $\boldsymbol{\Box(A \supset B) \supset (A \;\Box\!\!\rightarrow B)}$

Suppose that **A** could not have held without **B** holding too; then if **A** had held, **B** would also have held. In terms of possible worlds semantics for these operators along the lines of Lewis (1973) or Stalnaker (1968): if all **A** worlds are **B** worlds, then any closest **A** worlds are **B** worlds. More precisely, if all **A** worlds are **B** worlds, then either there are no **A** worlds or there is an **A** world such that any **A** world at least as close as it is to the actual world is a **B** world.

Second, the counterfactual conditional transmits possibility:

POSSIBILITY $(\mathbf{A} \mathrel{\Box\!\!\rightarrow} \mathbf{B}) \supset (\Diamond\mathbf{A} \supset \Diamond\mathbf{B})$

Suppose that if **A** had held, **B** would also have held; then if it is possible for **A** to hold, it is also possible for **B** to hold. In terms of worlds: if any closest **A** worlds are **B** worlds, and there are **A** worlds, then there are also **B** worlds. More precisely, if either there are no **A** worlds or there is an **A** world such that any **A** world at least as close as it is to the actual world is a **B** world, then if there is an **A** world there is also a **B** world.

Together, NECESSITY and POSSIBILITY sandwich the counterfactual conditional between two modal conditions. But they do not squeeze it very tight, for $\Diamond\mathbf{A} \supset \Diamond\mathbf{B}$ is much weaker than $\Box(\mathbf{A} \supset \mathbf{B})$: although the latter entails the former in any normal modal logic, the former is true and the latter false whenever **B** is possible without being a necessary consequence of **A**, for example when **A** and **B** are modally independent.

Although NECESSITY and POSSIBILITY determine no necessary and sufficient condition for the counterfactual conditional in terms of necessity and possibility, they yield necessary and sufficient conditions for necessity and possibility in terms of the counterfactual conditional.

We argue thus. Let $\perp$ be a contradiction. As a special case of NECESSITY:

(11) $\boldsymbol{\Box(\neg A \supset \perp) \supset (\neg A \mathrel{\Box\!\!\rightarrow} \perp)}$

By elementary (normal) modal logic, since a truth-functional consequence of something necessary is itself necessary:

(12) $\boldsymbol{\Box A \supset \Box(\neg A \supset \perp)}$

From (11) and (12) by transitivity of the material conditional:

(13) $\boldsymbol{\Box A \supset (\neg A \mathrel{\Box\!\!\rightarrow} \perp)}$

Similarly, as a special case of POSSIBILITY:

(14) $\boldsymbol{(\neg A \mathrel{\Box\!\!\rightarrow} \perp) \supset (\Diamond\neg A \supset \Diamond\perp)}$

By elementary (normal) modal logic, since the possibility of a contradiction is itself inconsistent, and necessity is the dual of possibility (being necessary is equivalent to having an impossible negation):

(15) $\mathbf{(\Diamond\neg A \supset \Diamond\bot) \supset \Box A}$

From (14) and (15) by transitivity:

(16) $\mathbf{(\neg A \mathrel{\Box\!\!\rightarrow} \bot) \supset \Box A}$

Putting (13) and (16) together:

(17) $\mathbf{\Box A \equiv (\neg A \mathrel{\Box\!\!\rightarrow} \bot)}$

The necessary is that whose negation counterfactually implies a contradiction. Since possibility is the dual of necessity (being possible is equivalent to having an unnecessary negation), (17) yields a corresponding necessary and sufficient condition for possibility, once a double negation in the antecedent of the counterfactual has been eliminated.

(18) $\mathbf{\Diamond A \equiv \neg(A \mathrel{\Box\!\!\rightarrow} \bot)}$

The impossible is that which counterfactually implies a contradiction; the possible is that which does not. In (17) and (18), the difference between necessity and possibility lies simply in the scope of negation.

Without assuming a specific framework for the semantics of counterfactuals (in particular, that of possible worlds), we can give a simple semantic rationale for (17) and (18), based on the idea of vacuous truth. That some true counterfactuals have impossible antecedents is clear, for otherwise $\mathbf{A} \Box\!\!\rightarrow \mathbf{A}$ would fail when **A** was impossible. Make two generally accepted assumptions about the distinction between vacuous and non-vacuous truth: (a) $\mathbf{B} \Box\!\!\rightarrow \mathbf{C}$ is vacuously true if and only if **B** is impossible (this could be regarded as a definition of 'vacuously' for counterfactuals); (b) $\mathbf{B} \Box\!\!\rightarrow \mathbf{C}$ is non-vacuously true only if **C** is possible. The truth of (17) and (18) follows, given normal modal reasoning. If $\Box\mathbf{A}$ is true, then $\neg\mathbf{A}$ is impossible, so by (a) $\neg\mathbf{A} \Box\!\!\rightarrow \bot$ is vacuously true; conversely, if $\neg\mathbf{A} \Box\!\!\rightarrow \bot$ is true, then by (b) it is vacuously true, so by (a) $\neg\mathbf{A}$ is impossible, so $\Box\mathbf{A}$ is true. Similarly, if $\Diamond\mathbf{A}$ is true, then **A** is not impossible, so by (a) $\mathbf{A} \Box\!\!\rightarrow \neg\bot$ is not vacuously true, and by (b) not non-vacuously true, so $\neg(\mathbf{A} \Box\!\!\rightarrow \bot)$ is true; if $\Diamond\mathbf{A}$ is not true, then **A** is impossible, so by (a) $\mathbf{A} \Box\!\!\rightarrow \neg\bot$ is vacuously true, so $\neg(\mathbf{A} \Box\!\!\rightarrow \bot)$ is not true.

Given that the equivalences (17) and (18) are logically true, metaphysically modal thinking is logically equivalent to a special case of counterfactual thinking, and the epistemology of the former is tantamount to a special case of the epistemology of the latter. Whoever has what it takes to understand the counterfactual conditional and the elementary logical auxiliaries $\neg$ and $\bot$ has what it takes to understand possibility and necessity operators.

The definability of necessity and possibility in terms of counterfactual conditionals was recognized long ago. It is easy to show from the closure and reflexivity principles for $\Box\!\!\rightarrow$ in §2 that $\mathbf{A} \Box\!\!\rightarrow \bot$ is logically equivalent to $\mathbf{A} \Box\!\!\rightarrow \neg\mathbf{A}$. Thus (17) and (18) generate two new equivalences:

(19) $\Box\mathbf{A} \equiv (\neg\mathbf{A} \Box\!\!\rightarrow \mathbf{A})$

(20) $\Diamond \mathbf{A} \equiv \neg(\mathbf{A} \mathrel{\Box\!\!\rightarrow} \neg \mathbf{A})$

The necessary is that which is counterfactually implied by its own negation; the possible is that which does not counterfactually imply its own negation. Stalnaker (1968) used (19) and (20) to define necessity and possibility, although his reading of the conditional (with a different notation) was not exclusively counterfactual. Lewis (1973a: 25) used (17) and (18) themselves to define necessity and possibility in terms of the counterfactual conditional. However, such definitions seem to have been treated as convenient notational economies, their potential philosophical significance unnoticed (Hill 2006 is a recent exception).

If we permit ourselves to quantify into sentence position ('propositional quantification'), we can formulate another pair of variants on (17) and (18) that may improve our feel for what is going on.[8] On elementary assumptions about the logic of such quantifiers and of the counterfactual conditional, $\neg\mathbf{A} \mathrel{\Box\!\!\rightarrow} \mathbf{A}$ is provably equivalent to $\forall \boldsymbol{p}\, (\boldsymbol{p} \mathrel{\Box\!\!\rightarrow} \mathbf{A})$: something is counterfactually implied by its negation if and only if it is counterfactually implied by everything. Thus (19) and (20) generate these equivalences too:

(21) $\Box \mathbf{A} \equiv \forall \boldsymbol{p}\, (\boldsymbol{p} \mathrel{\Box\!\!\rightarrow} \mathbf{A})$

(22) $\Diamond \mathbf{A} \equiv \exists \boldsymbol{p}\, \neg(\boldsymbol{p} \mathrel{\Box\!\!\rightarrow} \neg \mathbf{A})$

According to (21), something is necessary if and only if whatever were the case, it would still be the case (see also Lewis 1986: 23). That is a natural way of explaining informally what metaphysically necessity is. According to (22), something is possible if and only if it is not such that it would fail in every eventuality.

Since the right-hand sides of (17), (19) and (21) are not strictly synonymous with each other, given the differences in their semantic structure, they are not all strictly synonymous with $\Box$**A**. Similarly, since the right-hand sides of (18), (20) and (22) are not strictly synonymous with each other, they are not all strictly synonymous with $\Diamond$**A**. Indeed, we have no sufficient reason to regard any of the equivalences as strict synonymies. That detracts little from their philosophical significance, for failure of strict synonymy does not imply failure of logical equivalence. The main philosophical concerns about possibility and necessity apply equally to anything logically equivalent to possibility or necessity. A non-modal analogy: $\neg$**A** is logically equivalent to **A** $\supset \bot$, but presumably they are not strictly synonymous; nevertheless, once we have established that a creature can handle $\supset$ and $\bot$, we have established that it can handle something logically equivalent to negation, which answers the most interesting questions about its ability to handle negation. We should find the mutual equivalence of (17), (19) and (21), and of (18), (20) and (22) reassuring, for it shows the robustness of the modal notions definable from the counterfactual conditional, somewhat as the equivalence of the various proposed definitions of ‘computable function’ showed the robustness of that notion.

If we treat (17) and (18) like definitions of $\Box$ and $\Diamond$ for logical purposes, and assume some elementary principles of the logic of counterfactuals, then we can establish the main principles of elementary modal logic for $\Box$ and $\Diamond$. For example, we can show that what follows from necessary premises is itself necessary. Given that counterfactual conditionals obey modus ponens (or even weaker assumptions), we can show that what is necessary is the case. We can also check that the principles NECESSITY and POSSIBILITY, which we used to establish (17) and (18), do indeed hold under the latter characterizations of necessity and possibility. Under much stronger assumptions about the logic of the counterfactual conditional, we can also establish much stronger principles of modal logic, such as the S5 principle that what is possible is necessarily possible. Such connections extend to quantified modal logic. The logic of counterfactual conditionals smoothly generates the logic of the modal operators. Technical details are omitted here.

In particular, the proposed conception of modality makes quantification into the scope of modal operators tantamount to a special case of quantification into counterfactual contexts, as in (23) and (24):

(23) Everyone who would have benefited if the measure had passed voted for it.

(24) Where would the rock have landed if the bush had not been there?

Thus challenges to the intelligibility of claims of *de re* necessity are tantamount to challenges to the intelligibility of counterfactuals such as (23) and (24). But (23) and (24) are evidently intelligible.

Given (17) and (18), we should expect the epistemology of metaphysical modality to be a special case of the epistemology of counterfactuals. Far from being *sui generis*, the capacity to handle metaphysical modality is an 'accidental' byproduct of the cognitive mechanisms which provide our capacity to handle counterfactual conditionals. Since our capacity for modal thinking cannot be isolated from our capacity for ordinary thinking about the natural world, which involves counterfactual thinking, sceptics cannot excise metaphysical modality from our conceptual scheme without loss to ordinary thought about the natural world, for the former is implicit in the latter.

A useful comparison is with the relation between logical consequence and logical truth. Consider some agents who reason in simple ways about themselves and their environment, perhaps using rules of inference formalizable in a Gentzen-style natural deduction calculus, perhaps in some less sophisticated way. The practical value of their reasoning skill is that they can move from ordinary empirical premises to ordinary empirical conclusions in ways that always preserve truth, thereby extending their knowledge of mundane matters (see Schechter 2006 for relevant discussion). In doing so, they need never use logically true sentences. Nevertheless, the cognitive capacity that enables them to make these transitions between empirical sentences also enables them, as a special case, an 'accidental' byproduct, to deduce logical truths from the null set of premises. Highly artificial moves would be needed to block these bonus deductions; such *ad hoc* restrictions would come at the price of extra computational complexity for no practical gain. Likewise at the semantic level: the simplest compositional semantics that enables us to negate and conjoin empirical sentences also enables us to formulate logical truths and falsehoods, even if we have hitherto lacked any interest in doing so. By good fortune, everything is already in place for the logician to evaluate logical truths and falsehoods (at least in first-order logic, since it is complete). The philosopher's position with respect to metaphysical modality is not very different.

Discussions of the epistemology of modality often focus on imaginability or conceivability as a test of possibility while ignoring the role of the imagination in the assessment of mundane counterfactuals. In doing so, they omit the appropriate context for understanding the relation between modality and the imagination. For instance, scorn is easily poured on imagination as a test of possibility: it is imaginable but not possible that water does not contain oxygen, except in artificial senses of 'imaginable' that come apart from possibility in other ways, and so on. Imagination can be made to look cognitively worthless. Once we recall its fallible but vital role in evaluating counterfactual conditionals, we should be more open to the idea that it plays such a role in evaluating claims of possibility and necessity. At the very least, we cannot expect an adequate account of the role of imagination in the epistemology of modality if we lack an adequate account of its role in the epistemology of counterfactuals.

On the simplest version of the account in §2, we accept **A □→ B** when our counterfactual development of the supposition **A** generates **B**; we reject **A □→ B** when our counterfactual development of **A** fails to generate **B** (in a reasonable time). Thus, by (17), we accept □**A** when our counterfactual development of the supposition ¬**A** generates a contradiction; we reject □**A** when our counterfactual development of ¬**A** fails to generate a contradiction (in a reasonable time). Similarly, by (18), we accept ◊**A** when our counterfactual development of the supposition **A** fails to generate a contradiction (in a reasonable time); we reject ◊**A** when our counterfactual development of **A** generates a contradiction. Thus our fallible imaginative evaluation of counterfactuals has a conceivability test for possibility and an inconceivability test for impossibility as fallible special cases. Such conceivability and inconceivability will be subject to the same constraints, whatever they are, as counterfactual conditionals in general, concerning which parts of our background information are held fixed. If we know enough chemistry, our counterfactual development of the supposition that gold is the element with atomic number 79 will generate a contradiction. The reason is not simply that we know that gold is the element with atomic number 79, for we can and must vary some items of our knowledge under counterfactual suppositions. Rather, general constraints on the development of counterfactual suppositions require us to hold such constitutive facts fixed.

A nuanced account of our handling of counterfactuals is likely to predict that we are more reliable in evaluating some kinds than others. For example, we may well be more reliable in evaluating counterfactuals whose antecedents involve small departures from the actual world than in evaluating those whose antecedents involve much larger departures. We may be correspondingly more reliable in evaluating the possibility of everyday scenarios than of 'far-out' ones, and extra caution may be called for in the latter case. At the limit, actuality is often the best argument for possibility. But current philosophical practice already shows some sensitivity to such considerations. We may be more confident of the possibility of more or less realistic thought experiments in epistemology and moral philosophy than of more radically strange ones in metaphysics. More explicit consideration of the link between modal thought and counterfactual thought may lead to further refinements of our practice. But the use of imagination to evaluate philosophical claims of possibility and necessity is not illegitimate in principle, any more than is its use to evaluate mundane counterfactuals.

What does the envisaged assimilation of modality to counterfactual conditionals imply for the status of modal judgments as knowable *a priori* or only *a posteriori*? Some counterfactual conditions look like paradigms of *a priori* knowability: for example (7), whose consequent is a straightforward deductive consequence of its antecedent. Others look like paradigms of what can be known only *a posteriori*: for example, that if I had searched in my pocket five minutes ago I would have found a coin. But those are easy cases.

Standard discussions of the *a priori* distinguish between two roles that experience plays in cognition, one *evidential*, one *enabling*. Experience is held to play an evidential role in my visual knowledge that this shirt is green, but a merely enabling role in my knowledge that all green things are coloured: I needed it only to acquire the concepts *green* and *coloured*, without which I could not even raise the question whether all green things are coloured. Knowing *a priori* is supposed to be incompatible with an evidential role for experience, so my knowledge that this shirt is green is not *a priori*; but compatible with an enabling role for experience, so my knowledge that all green things are coloured can still be *a priori*. However, in our imagination-based knowledge of counterfactuals, experience can play a role that is neither strictly evidential nor purely enabling. For it can mould the ways in which we later imagine and judge, beyond what is needed to grasp the relevant concepts, without surviving as part of our total evidence.

Here is an example. I acquire the words 'inch' and 'centimetre' independently of each other. Through experience, I learn to make naked eye judgments of distances in inches or centimetres with moderate reliability. When things go well, such judgments amount to knowledge: *a posteriori* knowledge, of course. For example, I know *a* *posteriori* that two marks in front of me are at most two inches apart. Now I deploy the same faculty off-line to make a counterfactual judgment:

(25) If these marks had been at least nine inches apart, they would have been at least nineteen centimetres apart.

In judging (25), I do not use a conversion ratio between inches and centimetres to make a calculation. In the example I know no such ratio. Rather, I visually imagine the two marks nine inches apart, and use my ability to judge distances in centimetres visually off-line to judge under the counterfactual supposition that the marks are at least nineteen centimetres apart. With this large margin for error, my judgment is reliable. Thus I know (25). Do I know it *a priori* or *a posteriori*? Experience plays no direct evidential role in my judgment. I do not consciously or unconsciously recall memories of distances encountered in perception, nor do I deduce (25) from general principles that I have inductively or abductively gathered from experience: §2 noted obstacles to assimilating counterfactual thinking to reasoning. Nevertheless, the causal role of past experience in my judgment of (25) far exceeds enabling me to grasp the concepts in (25). Someone could easily have enough experience to understand (25) without being reliable enough in their judgments of distance to know (25).

If we classify my knowledge of (25) in the envisaged circumstances as *a priori*, because experience plays no strictly evidential role, the danger is that far too much will count as *a priori*. Experience can mould my judgment in many ways without playing a direct evidential role. But if we classify my knowledge of (25) as *a posteriori*, because experience plays more than a purely enabling role, that may apply to many philosophically significant modal judgments too. Of course, Kripke has argued strongly for a category of necessary truths knowable only *a posteriori*, such as ‘Gold is the element with atomic number 79’; ‘It is necessary that gold is the element with atomic number 79’ would then be knowable only *a posteriori* too. The present suggestion is intended far more widely than that. For example:

(26) It is necessary that whoever knows something believes it.

(27) If Mary knew that it was raining, she would believe that it was raining.

Knowledge of truths such as (26) and (27) is usually regarded as *a priori*, even by those who accept the category of the necessary *a posteriori*. The experiences through which we learned to distinguish in practice between belief and non-belief and between knowledge and ignorance play no strictly evidential role in our knowledge of (26) and (27). Nevertheless, their role may be more than purely enabling. Many philosophers, native speakers of English, have denied (26) (Shope 1983: 171-192 has a critical survey). They are not usually or plausibly accused of failing to understand the words ‘know’ and ‘believe’. Why should not subtle differences between two courses of experience, each of which sufficed for coming to understand ‘know’ and ‘believe’, make for differences in how test cases are imagined, just large enough to tip honest judgments in opposite directions? Whether knowledge of (26) and (27) is available to one may thus be highly sensitive to personal circumstances.

If that picture is on the right lines, should we conclude that modal knowledge is *a posteriori*? Not if that suggests that (26) and (27) are inductive or abductive conclusions from perceptual data. In such cases, the question '*A priori* or *a posteriori*?' is too crude to be of much epistemological use. The point is not that we cannot draw a line somewhere with traditional paradigms of the *a priori* on one side and traditional paradigms of the *a posteriori* on the other. Surely we can; the point is that doing so yields little insight. The distinction is handy enough for a rough initial description of epistemic phenomena; it is out of place in a deeper theoretical analysis, because it obscures more significant epistemic patterns.[9]

§4. It is time to consider objections to the preceding account.

Objection: Knowledge of counterfactuals cannot explain modal knowledge, because the former depends on the latter. More specifically, in developing a counterfactual supposition, we make free use of what we take to be necessary truths, but not of what we take to be contingent truths. Thus we rely on a prior stock of modal knowledge or belief. The principle NECESSITY above illustrates how we do this.

Reply: Once we take something to be a necessary truth, of course we can use it in developing further counterfactual suppositions. But that does nothing to show that we have any special cognitive capacity to handle modality independent of our general cognitive capacity to handle counterfactual conditionals. If we start only with the latter, just as envisaged above, it will generate knowledge of various modal truths, which can in turn be used to develop further counterfactual suppositions, in a recursive process. For example, we need not judge that it is metaphysically necessary that gold is the element with atomic number 79 *before* invoking the proposition that gold is the element with atomic number 79 in the development of a counterfactual supposition. Rather, projecting constitutive matters such as atomic numbers into counterfactual suppositions is part of our general way of assessing counterfactuals. The judgment of metaphysical necessity originates as the output of a procedure of that kind; it is not an independently generated input.

Objection: The account associates metaphysical modality with counterfactual conditionals of a very peculiar kind: in the case of (17) and (18), those with an explicit contradiction as their consequent. Why should a capacity to handle ordinary counterfactuals confer a capacity to handle such peculiar ones too?

Reply: That is like asking why a capacity to handle inferences between complex empirical sentences should confer a capacity to handle inferences involving logical truths and falsehoods too. There is no easy way to have the former without the latter. More specifically, developing a counterfactual supposition includes reasoning from it, and we cannot always tell in advance when such reasoning will yield a contradiction (there are surprises in logic). The undecidability of logical truth for first-order logic implies that there is no total mechanical test for the consistency of first-order sentences. Thus the inconsistent ones cannot be sieved out in advance (consider 'In the next village there is a barber who shaves all and only those in that village who do not shave themselves'). Consequently, a general capacity to develop counterfactual suppositions must confer in particular the capacity to develop those which subsequently turn out inconsistent. Although the capacity may not be of uniform reliability, as already noted, the variation is primarily with the *antecedent* of the counterfactual (the supposition under development), not with its consequent (which is what is exceptional in (17) and (18)). In deductive inference, our reasoning to contradictions (as in proof by *reduction ad absurdum*) is not strikingly more or less reliable than the rest of our deductive reasoning.

Objection: The assumption about vacuous truth on which the account relies is wrong (Nolan 1997). For some counterpossibles (counterfactuals with metaphysically impossible antecedents) are false, such as (28), uttered by someone who mistakenly believes that he answered '13' to 'What is 5 + 7?'; in fact he answered '11':

(28) If 5 + 7 were 13 I would have got that sum right.

Thus, contrary to (17), $\Box\mathbf{A}$ may be true while $\neg\mathbf{A} \mathrel{\Box\!\!\rightarrow} \bot$ is false. In the argument for (17) in §3, the objectionable premise is NECESSITY. If some worlds are metaphysically impossible, and **A** is true at some of them but false at all metaphysically possible worlds, while **B** is false at all worlds whatsoever, then every metaphysically possible **A** world is a **B** world, but the closest **A** worlds are not **B** worlds.[10] Similar objections apply to the other purported equivalences (18)-(22).

Reply: If *all* counterpossibles were false, $\Diamond\mathbf{A}$ would be equivalent to $\mathbf{A} \mathrel{\Box\!\!\rightarrow} \mathbf{A}$, for the latter would still be true whenever **A** was possible; correspondingly, $\Box\mathbf{A}$ would be equivalent to the dual $\neg(\neg\mathbf{A} \mathrel{\Box\!\!\rightarrow} \neg\mathbf{A})$ and one could carry out the programme of §3 using the new equivalences. But that is presumably not what the objector has in mind. Rather, the idea is that the truth-value of a counterpossible can depend on its consequent, so that (28) is false while (29) is true:

(29) If 5 + 7 were 13 I would have got that sum wrong.

However, such examples are quite unpersuasive.

First, they tend to fall apart when thought through. For example, if 5 + 7 were 13 then 5 + 6 would be 12, and so (by another eleven steps) 0 would be 1, so if the number of right answers I gave were 0, the number of right answers I gave would be 1.

Second, there are general reasons to doubt the supposed intuitions on which such examples rely. We are used to working with possible antecedents, and given the possibility of $\mathbf{A}$, the incompatibility of $\mathbf{B}$ and $\mathbf{C}$ implies that $\mathbf{A} \mathrel{\Box\!\!\rightarrow} \mathbf{B}$ and $\mathbf{A} \mathrel{\Box\!\!\rightarrow} \mathbf{C}$ cannot both be true. Thus by over-projecting from familiar cases we may take the uncontentious (29) to be incompatible with (28). The logically unsophisticated make analogous errors in quantificational reasoning. Given the evident truth of ‘Every golden mountain is a mountain’, they think that ‘Every golden mountain is a valley’ is false, neglecting the case of vacuous truth. Since the logic and semantics of counterfactual conditionals is much less well understood, even the logically sophisticated may find similar errors tempting. Such errors may be compounded by a tendency to confuse negating a counterfactual conditional with negating its consequent, given the artificiality of the constructions needed to negate the whole conditional unambiguously (‘it is not the case that if …’). Thus the truth of $\mathbf{A} \mathrel{\Box\!\!\rightarrow} \neg\mathbf{B}$ (with $\mathbf{A}$ impossible) may be mistaken for the truth of $\neg(\mathbf{A} \mathrel{\Box\!\!\rightarrow} \mathbf{B})$ and therefore the falsity of $\mathbf{A} \mathrel{\Box\!\!\rightarrow} \mathbf{B}$.

Some objectors try to bolster their case by giving examples of mathematicians reasoning from an impossible supposition $\mathbf{A}$ (‘There are only finitely many prime numbers’) in order to reduce it to absurdity. Such arguments can be formulated using a counterfactual conditional, although they need not be. Certainly there will be points in the argument at which it is legitimate to assert $\mathbf{A} \mathbin{\Box\!\!\rightarrow} \mathbf{C}$ (in particular, $\mathbf{A} \mathbin{\Box\!\!\rightarrow} \mathbf{A}$) but illegitimate to assert $\mathbf{A} \mathbin{\Box\!\!\rightarrow} \neg\mathbf{C}$ (in particular, $\mathbf{A} \mathbin{\Box\!\!\rightarrow} \neg\mathbf{A}$). But of course that does not show that $\mathbf{A} \mathbin{\Box\!\!\rightarrow} \neg\mathbf{A}$ is false. At any point in a mathematical argument there are infinitely many truths that it is not legitimate to assert, because they have not yet been proved (Lewis 1986: 24-6 pragmatically explains away some purported examples of false counterfactuals with impossible antecedents).

We may also wonder what logic of counterfactuals the objectors envisage. If they reject elementary principles of the pure logic of counterfactual conditionals, that is an unattractive feature of their position. If they accept all those principles, then they are committed to operators characterized as in (17) and (18) that exhibit all the logical behaviour standardly expected of necessity and possibility. What is that modality, if not metaphysical modality?

A final problem for the objection is this. Here is a paradigm of the kind of counterpossible which the objector regards as false:

(30) If Hesperus had not been Phosphorus, Phosphorus would not have been Phosphorus.

Since Hesperus is Phosphorus, it is metaphysically impossible that Hesperus is not Phosphorus, by the necessity of identity. Nevertheless, the objectors are likely to insist that in imaginatively developing the counterfactual supposition that Hesperus is not Phosphorus, we are committed to the explicit denial of no logical truth, as in the consequent of (30). According to them, if we do our best for the antecedent, we can develop it into a logically coherent though metaphysically impossible scenario: it will exclude ‘Phosphorus is not Phosphorus’. But they will presumably accept this trivial instance of reflexivity:

(31) If Hesperus had not been Phosphorus, Hesperus would not have been Phosphorus.

In general, however, coreferential proper names are intersubstitutable in counterfactual contexts. For example, the argument from (32) and (33) to (34) is unproblematically valid:

(32) If the rocket had continued on that course, it would have hit Hesperus.

(33) Hesperus = Phosphorus.

(34) If the rocket had continued on that course, it would have hit Phosphorus.

Similarly, the argument from (31) and (33) to (30) should be valid. But (31) and (33) are uncontentiously true. If the objector concedes that (30) is true after all, then there should be an explanation of the felt resistance to it, compatible with its truth, and we may reasonably expect that explanation to generalize to other purported examples of false counterpossibles. On the other hand, if objectors reject (30), they must deny the validity of the argument from (31) and (33) to (30). Thus they are committed to the claim that counterfactual conditionals create opaque contexts for proper names (the same argument could be given for other singular terms, such as demonstratives). But that is highly implausible. (32) and (34) are materially equivalent because their antecedents and consequents concern the same objects, properties and relations: it matters not that different names are used, because the counterfactuals are not about such representational features. But then exactly the same applies to (30) and (31). Their antecedents and consequents too concern the same objects, properties and relations. That the antecedent of (30) and (31) is in fact metaphysically impossible does not radically alter their subject matter. The transparency of the counterfactual conditional construction concerns its general logical form, not the specific content of the antecedent.

Under scrutiny, the case for false counterpossibles looks feeble.

Objection: Counterfactuals are desperately vague and context-sensitive; equivalences such as (17) and (18) will infect $\Box$ and $\Diamond$, interpreted as metaphysical modalities, with all that vagueness and context-sensitivity.

Reply: Infection is not automatic. For instance, within a Lewis-Stalnaker framework, different readings or sharpenings of $\Box\!\!\rightarrow$ may differ on the similarity ordering of worlds while still agreeing on what worlds there are, so that the differences cancel out in the right-hand sides of (17) and (18). Whether a given supposition counterfactually implies a contradiction may be unclear to us; that does not imply that there is no right answer.

Objection: It has been argued that counterfactual conditionals lack truth-values (Edgington 2003, Bennett 2003: 252-6). If so, the assimilation of claims of metaphysical possibility and necessity to counterfactuals will deprive such claims of truth-values.

Reply: The issues are too complex to discuss properly here, but the readily intelligible occurrence of counterfactual conditionals embedded in the scope of other operators as in (23) and (24) is hard to make sense of without attributing truth-values to the embedded occurrences. Here is another example:

(35) Every field that would have been flooded if the dam had burst was ploughed.

(35) can itself be intelligibly embedded in more complex sentences in all the usual ways. In order to understand how such embeddings work, we must assign truth-conditions to (35); *ad hoc* treatments of a few particular embeddings are not enough. For (35) to have truth-conditions, 'field that would have been flooded if the dam had burst' must have application-conditions. Thus there must be a distinction between the fields to which 'would have been flooded if the dam had burst' applies and those to which it does not. But that is just to say that there must be a distinction between the values of '*x*' for which 'If the dam had burst, *x* would have been flooded' is true and those for which it is false. That it is somewhat obscure what the truth-conditions of counterfactual conditionals are, and that we sometimes make conflicting judgments about them, hardly shows that they do not exist.

§5. The counterfactual conditional is of course not the only construction in ordinary use that is closely related to metaphysical modality. Consider comments after a swiftly extinguished fire in an explosives factory:

(36) There could have been a huge explosion.

(37) There could easily have been a huge explosion.

The truth-value of both (36) (so interpreted) and (37) depends on the location of the fire, the precautions in place, and so on. The mere metaphysical possibility of a huge explosion is insufficient to verify either (36) (so interpreted) or (37). The restricted nature of the possibility is explicit in (37) with the word 'easily'; it is implicit in the context of (36).[11] To discover the truth-value of (36) or (37), we need background information. We may also need our imagination, in attempting to develop a feasible scenario in which there is a huge explosion. We use the same general cognitive faculties as we do in evaluating related counterfactual conditionals, such as (38):

(38) If the fire engine had arrived a minute later, there would have been a huge explosion.

Judgments of limited possibility such as (36) (interpreted as above) and (37) have a cognitive value for us similar to that of counterfactual conditionals such as (38).

Both (36) and (37) entail (39), although not vice versa:

(39) It is metaphysically possible that there was a huge explosion.

This is another way in which our ordinary cognitive capacities enable us to recognize that something non-actual is nevertheless metaphysically possible. But we cannot reason from the negation of (36) or of (37) to the negation of (39).

Can metaphysical possibility be understood as the limiting case of such more restricted forms of possibility? Perhaps, but we would need some account of what demarcates the relevant forms of possibility from irrelevant ones, such as epistemic possibility. It also needs to be explained how, from the starting-point of ordinary thought, we manage to single out the limiting case, metaphysical modality. The advantage of counterfactual conditionals is that they allow us to single out the limiting case simply by putting a contradiction in the consequent; contradictions can be formed in any language with conjunction and negation Anyway, the connections with restricted possibility and with counterfactual conditionals are not mutually exclusive, for they are not being interpreted as rival semantic analyses, but rather as different cases in which the cognitive mechanisms needed for one already provide for the other.

The epistemology of metaphysical modality requires no dedicated faculty of intuition. It is simply a special case of the epistemology of counterfactual thinking, a kind of thinking tightly integrated with our thinking about the spatio-temporal world. To deny that such thinking ever yields knowledge is to fall into an extravagant scepticism. Here as elsewhere, we can do philosophy on the basis of general cognitive capacities that are in no deep way peculiarly philosophical.

Notes

[*] An earlier version of this paper was presented at the 2005 Erfurt conference on philosophical knowledge. It developed out of Williamson 2005, especially 15-22, from which it inherits numerous debts; Williamson 2004 explains other aspects of the associated general account of philosophical knowledge. Further debts were acquired from discussion after the presentation of that paper as a Presidential Address to the Aristotelian Society and of related material in the Blackwell Brown lectures at Brown University, the Anders Wedberg lectures at Stockholm University (where Sören Häggqvist and Anna-Sara Malmgren commented on the relevant lectures), at the third meeting of the Portugese Society for Analytic Philosophy in Lisbon and at conferences and workshops at the Centre for Advanced Studies in the Norwegian Academy of Sciences in Oslo, the Australian National University, Rutgers University and the universities of Bristol, Munich and Rochester, at colloquia at the University of California at Los Angeles and the universities of Arizona, Bologna, Heidelberg, Leeds, Nottingham, Turin and Warwick, and in Oxford, although the ideas presented here were not always centre stage. Thanks to all the individuals who have helped improve this paper, and in particular to Ann-Sara Malmgren and Thomas Kroedel for many valuable discussions of its themes.

[1] The large empirical literature on the affective role of counterfactuals and its relation to learning from experience includes Kahneman and Tversky 1982, Roese and Olson 1993, 1995 and Byrne 2005.

[2] The phrase ‘does in fact show’ is read throughout as inside the scope of the counterfactual conditional or modal operator, but as rigid, like ‘actually shows’. See Williamson 2006 for relevant discussion.

[3] Matters become more complicated if **A** or **B** itself contains a counterfactual condition, as in ‘If she had murdered the man who would have inherited her money if she had died, she would have been sentenced to life imprisonment if she had been convicted’.

[4] See Goldman 1992: 24, discussed by Nichols, Stich, Leslie and Klein 1996: 53-59.

[5] The question is of course related to Berkeley’s claim that we cannot imagine an unseen object. For discussion see Williams 1966, Peacocke 1985 and Currie 1995: 36-37.

[6] A similar problem arises for what is sometimes called the Ramsey Test for conditionals, on which one simulates belief in the antecedent and asks whether one then believes the consequent. Goldman writes ‘When considering the truth value of “If X were the case, then Y would obtain,” a reasoner feigns a belief in X and reasons about Y under that pretense” (1992: 24). What Ramsey himself says is that when people ‘are fixing their degrees of belief in *q* given *p*’ they ‘are adding *p* hypothetically to their stock of knowledge and arguing on that basis about *q*’ (1978: 143), but he specifically warns that ‘the degree of belief in *q* given *p*’ does not mean the degree of belief ‘which the subject

would have in *q* if he knew *p*, or that which he ought to have' (1978: 82; variables interchanged). Of course, conditional probabilities bear more directly on indicative than on subjunctive conditionals.

[7] Lewis defends the assumption (1986: 26-31); Nozick rejects it to make the fourth condition in his analysis of knowledge non-trivial (1981: 176). Bennett also rejects it (2003: 239-40).

[8] This quantification into sentence position need not be understood substitutionally. In purely modal contexts it can be modeled as quantification over all sets of possible worlds, even if not all of them are intensions of sentences that form the supposed substitution class, although this modeling presumably fails for hyperintensional contexts such as epistemic ones. A more faithful semantics for it might use non-substitutional quantification into sentence position in the meta-language. Such subtleties are inessential for present purposes.

[9] This problem for the *a priori/a posteriori* distinction undermines arguments for the incompatibility of semantic externalism with our privileged access to our own mental states that appeal to the supposed absurdity of *a priori* knowledge of contingent features of the external environment (McKinsey 1991).

[10] Technically, NECESSITY fails on a semantics with similarity spheres for □→ that include some impossible worlds (inaccessible with respect to □). Conversely,

POSSIBILITY fails on a semantics with some possible worlds excluded from all similarity spheres (see Lewis 1986: 16 on universality). Inaccessible worlds seem not to threaten POSSIBILITY. For suppose that an **A** world w but no **B** world is accessible from a world v. Then if **A** □→ **B** holds at v on the usual semantics, there is an **A** world x such that every **A** world as close as x is to v is a **B** world. It follows that w is not as close as x is to v and that x is inaccessible from v, which contradicts the plausible assumption that any accessible world is at least as close as any inaccessible world.

[11] On easy possibility see Sainsbury 1997, Peacocke 1999: 310-28 and Williamson 2000: 123-30. On the idea that natural language modals such as ‘can’ and ‘must’ advert to contextually restricted ranges of possibilities see Kratzer 1977.

Bibliography

Anderson, A.R. 1951. 'A note on subjunctive and counterfactual conditionals', *Analysis* 12: 35-8.

Bennett, J. 2003. *A Philosophical Guide to Conditionals*. Oxford: Clarendon Press.

Blackburn, S. 1987. 'Morals and modals', in G. Macdonald and C. Wright, eds., *Fact, Science and Morality*. Oxford: Blackwell.

Byrne, R.M.J. 2005. *The Rational Imagination: How People Create Alternatives to Reality*. Cambridge, Mass.: MIT Press.

Collins, J., Hall, N., and Paul, L. A. 2004. *Causation and Counterfactuals*. Cambridge, Mass.: MIT Press.

Craig, E. 1985. 'Arithmetic and fact', in I. Hacking, ed., *Essays in Analysis*, Cambridge: Cambridge University Press.

Currie, G. 1995. 'Visual imagery as the simulation of vision', *Mind and Language* 10: 17-44.

Davies, M., and Stone, T., eds. 1995. *Mental Simulation: Evaluations and Applications*, Oxford: Blackwell.

Edgington, D. 2003. 'Counterfactuals and the benefit of hindsight', in P. Dowe and P. Noordhof, eds., *Causation and Counterfactuals*. London: Routledge.

Evans, J. St. B. T., and Over, D. E. 2004. *If*. Oxford: Oxford University Press.

Gendler, T. Szabó, and Hawthorne, J., eds. 2002. *Conceivability and Possibility*, Oxford: Clarendon Press.

Goldman, A. 1992. ‘Empathy, mind and morals’, *Proceedings and Addresses of the American Philosophical Association* 66/3: 17-41.

Goodman, N. 1955. *Fact, Fiction, and Forecast*. Cambridge, MA: Harvard University Press.

Harris, P. 2000. *The Work of the Imagination*. Oxford: Blackwell.

Hill, C. S. 2006. ‘Modality, modal epistemology, and the metaphysics of consciousness’, in S. Nichols (ed.), *The Architecture of the Imagination: New Essays on Pretense, Possibility and Fiction*. Oxford: Oxford University Press.

Jackson, F. 1977. ‘A causal theory of counterfactuals’, *Australasian Journal of Philosophy* 55: 3-21.

Kahneman, D., and Tversky, A. 1982. ‘The simulation heuristic’, in D. Kahneman, P. Slovic and A. Tversky, eds., *Judgement under Uncertainty*. Cambridge: Cambridge University Press.

Kaplan, D. 1989. ‘Demonstratives: An Essay on the Semantics, Logic, Metaphysics, and Epistemology of Demonstratives and Other Indexicals’ in J. Almog, J. Perry and H. Wettstein, eds, *Themes from Kaplan*. Oxford: Oxford University Press.

Kratzer, A. 1977. ‘What “must” and “can” must and can mean’, *Linguistics and Philosophy* 1: 337-355.

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

Lewis, D. 1973a. ‘Counterfactuals and comparative possibility’, *Journal of Philosophical Logic* 2: 418-46. Reprinted in his *Philosophical Papers*, vol. 2, Oxford: Oxford University Press, 1986, to which page numbers refer.

Lewis, D. 1973b. ‘Causation’, *Journal of Philosophy* 70: 556-67.

Lewis, D. 1979. ‘Counterfactual dependence and time’s arrow’, *Noûs* 13: 455-476.

Lewis, D. 1986. *Counterfactuals*, revised edn. Cambridge, Mass.: Harvard University Press.

McKinsey, M. 1991. ‘Anti-individualism and privileged access’, *Analysis* 51: 9-16.

Nichols, S., and Stich, S. P. 2003. *Mindreading: An Integrated Account of Pretence, Self -Awareness, and Understanding of Other Minds*. Oxford: Clarendon Press.

Nichols, S., Stich, S. P, Leslie, A., and Klein, D. 1996. ‘Varieties of off-line simulation’, in P. Carruthers and P. K. Smith, eds., *Theories of Theories of Mind*. Cambridge: Cambridge University Press.

Nolan, D. 1997. ‘Impossible worlds: a modest approach’, *Notre Dame Journal for Formal Logic* 38: 535-572.

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

Peacocke, C. 1985. ‘Imagination, experience and possibility’, in J. Foster and H. Robinson, eds., *Essays on Berkeley: A Tercentennial Celebration*. Oxford: Clarendon Press.

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

Ramsey, F. P. 1978. *Foundations: Essays in Philosophy, Logic, Mathematics and Economics*, ed. D. H. Mellor. London: Routledge & Kegan Paul.

Roese, N. J., and Olson, J. 1993. ‘The structure of counterfactual thought’, *Personality and Social Psychology Bulletin* 19: 312-19.

Roese, N. J., and Olson, J. 1995. ‘Functions of counterfactual thinking’, in N. J. Roese and J. M. Olson, eds., *What Might Have Been: The Social Psychology of Counterfactual Thinking*. Mahwah, NJ: Erlbaum.

Sainsbury, R. M. 1997. 'Easy possibilities', *Philosophy and Phenomenological Research*, 57: 907-19.

Schechter, J. 2006. 'Can evolution explain the reliability of our logical beliefs', typescript.

Shope, R. K. 1983. *The Analysis of Knowing: A Decade of Research*. Princeton: Princeton University Press.

Stalnaker, R. 1968. 'A theory of conditionals', in *American Philosophical Quarterly Monographs* 2 (*Studies in Logical Theory*): 98-112.

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

Williams, B. 1966. 'Imagination and the self', *Proceedings of the British Academy* 52: 105-124.

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

Williamson, T. 2004. 'Philosophical "intuitions" and skepticism about judgement', *Dialectica* 58: 109-153.

Williamson, T. 2005. 'Armchair philosophy, metaphysical modality and counterfactual thinking', *Proceedings of the Aristotelian Society* 105: 1-23.

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

Wright, C. 1989. 'Necessity, caution and scepticism', *Aristotelian Society* sup. 63: 203-38.