# Knowledge of Counterfactuals

TIMOTHY WILLIAMSON

Here are two claims:

(0I) If my enemies tried to murder me yesterday, they failed.

(0) If my enemies had tried to murder me yesterday, they would have failed.

In some sense that requires clarification, the antecedent of the indicative conditional (0I) supposes that my enemies *actually* tried to murder me, while the antecedent of the 'subjunctive' or 'counterfactual' conditional (0) supposes only that they tried to murder me in hypothetical circumstances without supposing those circumstances to be actual. I can easily know (0I) because I know that I am still alive. It is harder for me to know (0). Perhaps my enemies are clever and determined; my evidence may indicate that if they had tried, they would have succeeded. That I am still alive indicates that they did not try to murder me, not that they would have failed if they had tried. But (0) is not impossible to know. Perhaps, instead, I have bugged my enemies' discussions, and know that the murder plan they have ready for me depends on a false assumption about my whereabouts. Yet knowledge of such counterfactuals is puzzling. We cannot observe things that might have happened but didn't; nor can we observe their causes or effects.

Knowledge of counterfactuals has a special significance for philosophy. For many philosophical claims concern whether something that does not occur nevertheless could have occurred: for instance, time without space. In the jargon, they concern metaphysical possibility, impossibility and necessity. Our knowledge of these matters, such as it is, has grown out of our knowledge of far more mundane counterfactual matters, such as (0).

The aim of this essay is to sketch a picture of our ordinary knowledge of counterfactuals, and then to use it to raise a problem for the traditional philosophical dichotomy between *a priori* and *a posteriori* knowledge.[1]

[1] For the relation of the present account to knowledge of metaphysical modality see Williamson 2007, on which this paper draws.

doi:10.1017/S135824610900006X 

We can usefully start with a well-known example which proves the term 'counterfactual conditional' misleading. To adapt an example from Alan Ross Anderson (1951: 37), a doctor might say:

(1) If Jones had taken arsenic, he would have shown such-and-such symptoms.

We observe:

(2) Jones shows such-and-such symptoms.

Clearly, (1) and (2) can provide abductive evidence by inference to the best explanation for the antecedent of (1) (see Edgington 2003: 23–7 for more discussion):

(3) Jones took arsenic.

If further tests subsequently verify (3), 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. Rather, what the antecedent of (1) does not suppose is that Jones *actually* took arsenic. In what follows, we shall be just as concerned with conditional sentences such as (1) as with those whose premises are false, or believed to be so.

While (1) adds valuable empirical evidence to (2), the corresponding indicative conditional does not:

(1I) If Jones took arsenic, he shows such-and-such symptoms.

We can safely assent to (1I) just on the basis of inspecting Jones's corpse and observing (2), before hearing what the doctor has to say, simply because we can see that Jones *does* show such-and-such symptoms, whether or not he took arsenic. Informally, (1) is more useful than (1I) because (1), unlike (1I), depends on a comparison between independently specified terms, the symptoms Jones would have shown if he had taken arsenic and the symptoms he does in fact show. 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, given (2). 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 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:

(4) If Jones had taken arsenic, he would not have shown such-and-such symptoms.

From (2) and (4) we can deduce (5), the negation of its antecedent, for a counterfactual conditionals with a true antecedent and a false consequent is false:

(5) Jones did not take arsenic.

The indicative conditional corresponding to (4) is:

(4I) If Jones took arsenic, he does not show such-and-such symptoms.

Since we can clearly see that Jones does show such-and-such symptoms, to assert (4I) is like saying 'If Jones took arsenic, pigs can fly'. Although a very confident doctor might assert (4I), on the grounds that Jones certainly did not take arsenic, that certainty may in turn be based on confidence in (4), and therefore on the comparison of actual and counterfactual situations.

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

(6) 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.[2]

Evidently, counterfactual conditionals give clues to causal connections. This point does not commit one to any ambitious programme of analysing causality in terms of counterfactual conditionals (Lewis 1973, 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

[2] 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.



https://doi.org/10.1017/S135824610900006X Published online by Cambridge University Press

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.

In discussing the epistemology of counterfactuals, I assume no particular theory of the semantics of the counterfactual conditional. In particular, I do not assume the Stalnaker-Lewis approach, on which a counterfactual conditional statement is true in a given possible world if and only if either the consequent is true in the closest possible world or worlds to the given one or (the vacuous case) the antecedent is false in all possible worlds, where closeness is measured by similarity in certain respects (Stalnaker 1968, Lewis 1979, 1986). However, the Stalnaker-Lewis approach will occasionally be used 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 into the lake at the bottom. Under suitable background conditions, you thereby come to know this counterfactual:

(7) 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 (7) even without performing such experiments. Logically, 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 (1954). 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.



https://doi.org/10.1017/S135824610900006X Published online by Cambridge University Press

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

(8) 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 (8) on that basis. Similarly, it may be suggested, we can assert (7) 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 a conclusion from some premises, we can derive the counterfactual conditional that if a specified state of affairs had obtained the conclusion would have held from counterfactual conditionals to the effect that if the state of affairs had obtained the premises would have held; in other words, the counterfactual consequences of a supposition are closed under logical consequence. With the trivial principle that if a state of affairs had obtained it would have obtained, it follows that, given a derivation of a conclusion from the supposition that a specified state of affairs obtains alone, we need no extra premises to derive the counterfactual conditional that if the state of affairs had obtained the conclusion would have held.

We cannot automatically extend the closure rule to the case where there are auxiliary premises. For example, from the premises 'She won the match' and 'She broke her leg' we can trivially derive the conclusion 'She won the match', but we cannot legitimately move from that to deriving the counterfactual conclusion 'If she had broken her leg she would have won the match' from the premise 'She won the match', since the latter may be true when the former is false. Auxiliary premises cannot always be copied into the scope of counterfactual suppositions (this is 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. Two will be discussed here.

First, the putative reasoner may lack general-purpose cognitive access to the auxiliary premises of the putative reasoning. In particular, the folk physics needed to derive the consequents of counterfactuals such as (7) from their antecedents may be stored in the form of some analogue mechanism, perhaps embodied in a connectionist



https://doi.org/10.1017/S135824610900006X Published online by Cambridge University Press

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 formulate its negation. More generally, the supposed premises may not be stored in a form that permits the normal range of inferential interactions with other beliefs, even at an unconscious level. This strains the analogy with explicit reasoning.

The other problem is epistemological. Normally, someone who believes a conclusion on the sole basis of inference from some premises knows the conclusion only if they know the premises. This principle must be applied with care, for often a thinker is aware of several inferential routes from different sets of premises to the same conclusion. For example, you believe that *a* and *b* are *F*; you deduce that something is *F*. If you know that *a* is *F*, you may thereby come to know that something is *F*, even if your belief that *b* is *F* is false, and so not knowledge. Similarly, you may believe more premises than you need to draw an inductive conclusion. The principle applies only to essential premises, those that figure in all the inferences on which the relevant belief in the conclusion is based. However, folk physics is an essential standing background premise of the supposed inferences from antecedents to consequents of counterfactuals like (7), as usually conceived, so the epistemological maxim applies. Folk physics in this sense is a theory whose content includes the general principles by which expectations of motion, constancy and the like are formed on-line in real time; it is no mere collection of memories of particular past incidents. But then presumably it is strictly speaking false: although many of its predictions are useful approximations, they are inaccurate in some circumstances; knowledge of the true laws of motion is not already wired into our brains, otherwise physics could be reduced to psychology. Since folk physics is false, 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



https://doi.org/10.1017/S135824610900006X Published online by Cambridge University Press

folk physics as a locally but not globally reliable method of belief formation, not as a premise.

If folk theories are methods of belief formation rather than specific beliefs, can they be treated as patterns of inference, for example from beliefs about the present to beliefs about the future? Represented as a universal generalization, a non-deductive pattern of inference such as abduction is represented as a falsehood, for the relevantly best explanations are not always correct. Nevertheless, we can acquire knowledge abductively because we do not rely on every abduction in relying on one; we sometimes rely on a locally truth-preserving abduction, even though abduction is not globally truth-preserving. The trouble with replacing a pattern of inference by a universal generalization is that it has us rely on all instances of the pattern simultaneously, by relying on the generalization. Even if the universal generalization is replaced by a statement of general tendencies, what we are relying on in a particular case is still inappropriately globalized. Epistemologically, folk 'theories' seem to function more like patterns of inference than like general premises. That conception also solves the earlier problem about the inapplicability of logical operators to folk 'theories', since patterns of inference cannot themselves be negated or made the antecedents of conditionals (although claims of their validity can).

Once such a liberal conception of patterns of inference is allowed, calling a process of belief formation 'inferential' is no longer very informative. Just about any process with a set of beliefs (or suppositions) as input and an expanded set of beliefs (or suppositions) as output counts as 'inferential'. Can we say something more informative about the imaginative exercises by which we judge counterfactuals like (7), whether or not we count them as 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. For example, the cognitive faculties that would be run on-line to evaluate 'She broke her leg' and 'She won the match' as free-standing sentences are run off-line in the evaluation of the counterfactual conditional 'If she had broken her leg she would have won the match'.[3] This suggests that the cognition has a roughly compositional

[3] Matters become more complicated if the antecedent or consequent 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', but the underlying principles are the same.

structure. Our capacity to handle a counterfactual conditional embeds our capacities to handle its antecedent and consequent separately, 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 'If there had been infinitely many stars there would have been infinitely many stars' is true even if we have no idea how to determine whether 'There are infinitely many stars' 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 (7).

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



https://doi.org/10.1017/S135824610900006X Published online by Cambridge University Press

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?

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 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:

(9) 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 (9) 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

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

the observer as perfectly transparent.[5] Nor, when we consider (9), 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 is better not to regard the simulation as referring to anything specifically *mental* 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).

The use of expectation-forming capacities may in effect impose a partial solution to Goodman's problem of cotenability, since they do not operate on information about what happened after the time treated as present. In this respect indicative conditionals are evaluated differently: if I had climbed a mountain yesterday I would remember it today, but if I did climb a mountain yesterday I do not remember it

[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 pretence" (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). Conditional probabilities bear more directly on indicative than on counterfactual conditionals.



https://doi.org/10.1017/S135824610900006X Published online by Cambridge University Press

today. The known fact that I do not remember climbing a mountain yesterday is retained under the indicative but not the counterfactual supposition.

Our off-line use of expectation-forming capacities to unroll a counterfactual history from the imagined initial conditions does not explain why we imagine the initial conditions in one way rather than another – for instance, why we do not imagine a wall in place of the bush. Very often, no alternative occurs to us, but that does not mean that the way we go adds nothing to the given antecedent. We seem to have a prereflective tendency to minimum alteration in imagining counterfactual alternatives to actuality, reminiscent of the role that similarity between possible worlds plays in the Lewis-Stalnaker semantics.

Of course, not all counterfactual conditionals can be evaluated by the rolling back method, since the antecedent need not concern a particular 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. For it can be shown that any sentence whatsoever is equivalent to a counterfactual conditional, for example, to one with that sentence as the consequent and a tautology as the antecedent. Thus, *modulo* the recognition of this elementary equivalence, any cognitive work needed to evaluate the original sentence is also needed to evaluate the counterfactual conditional.

We can schematize the process of evaluating a counterfactual conditional thus: one imaginatively supposes the antecedent and develops the supposition, adding further judgments within the supposition by reasoning, off-line predictive mechanisms and other off-line judgments. All of one's background knowledge and belief is available from within the scope of the supposition as a description of one's actual circumstances for the purposes of comparison with the counterfactual circumstances (in this respect the development differs from that of the antecedent of an indicative conditional). Some but not all of one's background knowledge and belief is also available within the scope of the supposition as a description of the counterfactual circumstances, according to complex criteria (the problem of cotenability). To a first approximation: one asserts the counterfactual conditional if and only if the development eventually leads one to add the consequent.

An over-simplification in that account is that one develops the initial supposition only once. In fact, if one finds various different ways of imagining the antecedent equally good, one may try developing several of them, to test whether they all yield the consequent. For example, if in considering (9) one initially imagines a palm tree, one does 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 one knows that one can equally easily imagine a fir tree. One repeats the thought experiment. Robustness in the result under such minor perturbations supports a higher degree of confidence.

What happens if the counterfactual development of the antecedent does not robustly yield the consequent? We do not always deny the counterfactual, for several reasons. First, 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, one must in effect judge that if the consequent were ever going to emerge it would have done so by now. For example, one may have been smoothly fleshing

out a scenario incompatible with the consequent with no hint of difficulty. Second, even if one is confident that the consequent will not robustly emerge from the development, one may suspect that the reason is one's ignorance of relevant background conditions rather than the lack of a counterfactual connection between the antecedent and the consequent ('If I were to follow that path, it would lead me out of the forest'). Thus one may remain agnostic over the counterfactual.

The case for denying the counterfactual is usually strongest when the counterfactual development of the antecedent robustly yields the negation of the consequent. Then one asserts the opposite counterfactual, with the same antecedent and the negated consequent. The default is to deny a counterfactual if one asserts the opposite counterfactual, for example moving from 'If she had broken her leg she would have failed to win the match' to 'It is not the case that if she had broken her leg she would have won the match'. The move is defeasible; sometimes one must accept opposite counterfactuals together. For example, deductive closure generates both 'If she had both won and failed to win the match she would have won the match' and 'If she had both won and failed to win the match she would have failed to win the match'. Normally, if the counterfactual development of the antecedent robustly yields the negation of the consequent and robustly fails to yield the consequent itself then one denies the original counterfactual, but even this connection is defeasible, since one may still suspect that the original consequent (as well as its negation) would robustly emerge given more complex reasoning or further background information.

Sometimes a counterfactual antecedent is manifestly neutral between contradictory consequents: consider 'If the coin had been tossed it would have come up heads' and 'If the coin had been tossed it would have come up tails'. In such cases one will clearly never be in a position to assert one conditional, and thus will never be in a position to use it as a basis for denying the opposite conditional.

The epistemological asymmetry between asserting and denying a counterfactual conditional resembles an epistemological asymmetry in practice between asserting and denying many existential claims. If I find snakes in Iceland, without too much fuss I can assert that there are snakes in Iceland. If I fail to find snakes in Iceland, I cannot deny that there are snakes in Iceland without some implicit or explicit assessment of the thoroughness of my search: if there were snakes in Iceland, would I have found some by now? But we are capable of making such assessments, and sometimes are in a position to deny such existential claims. Similarly, if I find a counterfactual

connection between the antecedent and the consequent (my counterfactual development of the former robustly yields the latter) without too much fuss I can assert the counterfactual. If I fail to find a counterfactual connection between the antecedent and the consequent (my counterfactual development of the former does not robustly yield the latter), I cannot deny the counterfactual without some implicit or explicit assessment of the thoroughness of my search: if there were a counterfactual connection, would I have found it by now? But we are capable of making such assessments, and sometimes are in a position to deny counterfactual conditionals.

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.

Some counterfactual conditions look like paradigms of *a priori* knowability: for example (8), 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, or at least with an evidential role for sense experience, so my knowledge that this shirt is green is not *a priori*. By contrast, knowing *a priori* is supposed to be 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, sense experience can play a role that is neither strictly evidential nor purely enabling. For, even without surviving as part of our total evidence, it can mould our habits of imagination and judgment in ways that go far beyond a merely enabling role.

Here is an example. I acquire the words 'inch' and 'centimetre' independently of each other. Through sense 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:

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

In judging (10), 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 two marks nine inches apart, and use my ability to judge distances in centimetres visually off-line to judge under the counterfactual supposition that they are at least nineteen centimetres apart. With this large margin for error, my judgment is reliable. Thus I know (10). Do I know it *a priori* or *a posteriori*? Sense 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 (10) from general premises I have inductively or abductively gathered from experience: we noted above obstacles to assimilating such patterns of counterfactual judgment to the use of general premises. Nevertheless, the causal role of past sense experience in my judgment of (10) far exceeds enabling me to grasp the concepts relevant to (10). Someone could easily have enough sense experience to understand (10) without being reliable enough in their judgments of distance to know (10). Nor is the role of past experience in the judgment of (10) purely enabling in some other way, for example by acquainting me with a logical argument for (10). It is more directly implicated than that. Whether my belief in (10) constitutes knowledge is highly sensitive to the accuracy or otherwise of the empirical information about lengths (in each unit) on which I relied when calibrating my judgments of length (in each unit). I know (10) only if my off-line application of the concepts of an inch and a centimetre was sufficiently skilful. My possession of the appropriate skills depends constitutively, not just causally, on past experience for the calibration of my judgments of length in those units. If the calibration is correct by a lucky accident, despite massive errors in the relevant past beliefs about length, I lack the required skill.[7]

If we knew counterfactual conditionals by purely *a priori* inference from the antecedent and background premises to the conclusion, our

[7] Yablo 2002 has a related discussion of the concept *oval*.

knowledge might count as *a priori* if we knew all the background premises *a priori*, and otherwise as *a posteriori*. However, it was argued above that if the process is inferential at all, the relevant inferences are themselves of just the kind for which past experience plays a role that is neither purely enabling nor strictly evidential, so the inferential picture does not resolve the issue.

If we classify my knowledge of (10) in the envisaged circumstances as *a priori*, because sense experience plays no strictly evidential role, the danger is that far too much will count as *a priori*. Long-forgotten experience can mould my judgment in many ways without playing a direct evidential role, for example by conditioning me into patterns of expectation which are called on in my assessment of ordinary counterfactual conditionals. But if we classify my knowledge of (10) as *a posteriori*, because experience plays more than a purely enabling role, that may apply to many philosophically significant judgments too. For example:

(11) If you had been morally obliged to give the money, you would have been able to give it.

If we know (11), our way of knowing it is similar to our way of knowing (10). Knowledge of truths like (11) 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 the obligatory and the non-obligatory and between ability and inability play no strictly evidential role in our knowledge of (11). Nevertheless, their role may be more than purely enabling. Why should not subtle differences between two courses of experience, each of which sufficed for coming to understand (11), make for differences in how test cases are processed, just large enough to tip honest judgments in opposite directions? Whether knowledge of (11) is available to one may thus be highly sensitive to personal circumstances. Such individual differences in the skill with which concepts are applied depend constitutively, not just causally, on past experience, for the skillfulness of a performance depends constitutively on its causal origins.

In a similar way, past experience of spatial and temporal properties may play a role in skilful mathematical 'intuition' that is not directly evidential but far exceeds what is needed to acquire the relevant mathematical concepts. The role may be more than heuristic, concerning the context of justification as well as the context of discovery. Even the combinatorial skills required for competent assessment of standard set-theoretic axioms may involve off-line applications of perceptual and motor skills, whose capacity to generate knowledge

constitutively depends on their honing through past experience that plays no evidential role in the assessment of the axioms.

If the preceding picture is on the right lines, should we conclude that modal knowledge is *a posteriori*? Not if that suggests that (11) is an inductive or abductive conclusion 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. We may acknowledge an extensive category of *armchair knowledge*, in the sense of knowledge in which experience plays no strictly evidential role, while remembering that such knowledge may not fit the stereotype of the *a priori*, because the contribution of experience was far more than enabling. For example, it should be no surprise if we turn out to have armchair knowledge of truths about the external environment.[8]

*New College, Oxford*

## Bibliography

Anderson, A. R. (1951) "A note on subjunctive and counterfactual conditionals", *Analysis* **12**, 35–8.

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

Chalmers, D. J. (2006) "The foundations of two-dimensional semantics", in M. García-Carpintero and J. Macià, eds., *Two-Dimensional Semantics*, Oxford: Clarendon Press.

Collins, J., Hall, N., and Paul, L. A. (2004) *Causation and Counterfactuals*. Cambridge, MA: M.I.T. Press.

[8] 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). It also renders problematic attempts to explain the first dimension of two-dimensional semantics in terms of *a priori* knowability, as in Chalmers 2006. Substituting talk of rational reflection for talk of the *a priori* does not help, since it raises parallel questions.

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: Evaluation and Applications*. Oxford: Blackwell.
Edgington, D. (2003) "Counterfactuals and the Benefit of Hindsight", *Causation and Counterfactuals*, (eds.) P. Dowe and P. Noordhof. London: Routledge.
Evans, J. St. B. T., and Over, D. E. (2004) *If*. Oxford: Oxford University Press.
Goldman, A. (1992) "Empathy, mind and morals", *Proceedings and Addresses of the American Philosophical Association* **66/3**, 17–41.
Goodman, N. (1954) *Fact, Fiction and Forecast*. London: Athlone Press.
Harris, P. (2000) *The Work of the Imagination*. Oxford: Blackwell.
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 *Judgement under Uncertainty*, (eds.) D. Kahneman, P. Slovic and A. Tversky. Cambridge: Cambridge University Press.
Lewis, D. K. (1973) "Causation", *Journal of Philosophy* **70**, 556–67.
———— (1979) "Counterfactual dependence and time's arrow", *Noûs* **13**, 455–476.
———— (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 *Theories of Theories of Mind*, (eds.) P. Carruthers and P. K. Smith. Cambridge, Cambridge University Press.
Peacocke, C. A. B. (1985) "Imagination, experience and possibility", in *Essays on Berkeley: A Tercentennial Celebration*, (eds.) J. Foster and H. Robinson. 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.

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

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

Williams, B. (1966) "Imagination and the self", *Proceedings of the British Academy* **52**, 105–124.

Williamson, T. (2007) *The Philosophy of Philosophy*. Oxford: Blackwell.

Yablo, S. (2002) "Coulda, woulda, shoulda", in *Conceivability and Possibility*, (eds.) T. S. Gendler and J. Hawthorne. Oxford: Clarendon Press.