

## A Dilemma for any Theory of Knowledge

Trenton Merricks

*American Philosophical Quarterly* (1995): 279-284

Either present knowledge depends on things that have not yet happened or one can have very little knowledge about the future. For two worlds ( $W$  and  $W^*$ ) that share the same events up to a time  $t$ , the same belief at  $t$  can be accidentally true in  $W^*$  and therefore not known, but both true and known in  $W$ , depending on what happens in the future in  $W$  and  $W^*$ . If truth values are denied for future contingent propositions, then this threatens the ability to have knowledge about the future. Otherwise, knowledge at  $t$  depends on what happens after  $t$ , with significant implications for the theory of knowledge.

I

Two worlds "share an initial segment" up to, and including, a time  $t$  only if the same individuals exist (at the same times), and the same events occur in the same ways (at the same times), in each world up to and including time  $t$ . Worlds that share an initial segment up to time  $t$  can differ in some ways at or before  $t$ , but they can differ only with respect to "soft facts", not "hard facts".<sup>1</sup> For instance, two worlds could share an initial segment up to and including time  $t$ , but differ, at  $t$ , in that one and not the other is such that it will rain one hour after  $t$ . Those worlds could not, however, differ with respect to whether it is raining at (or before)  $t$ . None of this, of course, comes anywhere near a helpful definition of "initial segment" or "hard fact" or "soft fact". Fortunately, nothing more than an intuitive grasp of these notions is necessary to understand the arguments that follow.

Consider the following principle:

(1) If two worlds share an initial segment up to, and including time  $t$ , and if at  $t$   $S$  truly believes that  $p$  in both worlds, then at  $t$   $S$  knows that  $p$  in one world if and only if  $S$  knows that  $p$  in the other.<sup>2</sup>

(1) asserts that if  $S$  knows that  $p$  in some world, then  $S$  knows that  $p$  in every other world that is just like that world up to, and including, the time at which  $S$  truly believes that  $p$ . To put the same point a different way, (1) states that if  $S$  (believes  $p$  but) fails to know that  $p$  in some world in which  $p$  is true, then in all possible worlds exactly like that world up to, and including, the time of  $S$ 's truly believing that  $p$ ,  $S$  fails to know that  $p$ . In order for (1) to avoid triviality, we cannot assume that the worlds are similar at time  $t$  with respect to  $S$ 's knowing that  $p$ .

Suppose that I believe there will be no dinosaur in the hallway outside of my office in one second (I do believe this; indeed, I feel certain of it). Let us add that this belief is true. Further, suppose that I am impressed by the maxim that one ought always to increase the number of one's true beliefs. I therefore consider, and believe, the following disjunction:

(P) There will be no dinosaur in the hallway in one second or Brown is in Barcelona.

Let us also add that although I do not know that Brown is in Barcelona (I have no idea as to where Brown might be), Brown is in Barcelona.

It is broadly logically possible, given all that has occurred in the world so far, that a dinosaur appear in the hallway in one second. In other words, there is a possible world  $W^*$  which shares an initial segment with this world ( $W$ ) up to, and including, the present time ( $t$ ), but in which a dinosaur miraculously appears in the hallway in one second. In  $W^*$ , as in  $W$ , I believe that  $p$ . Add that in both  $W$  and  $W^*$  Brown is (at  $t$ ) in Barcelona, so that  $p$  is true in both worlds. But in  $W^*$  my belief that  $p$  is accidentally true. It is only because Brown happens (unbeknownst to me) to be in Barcelona that  $p$  is true. Since I cannot know what is only accidentally true, I do not know that  $p$  in  $W^*$ .

Nothing here turns on controversial claims about the nature of accidental truth. For our purposes,  $S$ 's belief that  $p$  is accidentally true just in case  $p$  is both true and obviously not known by  $S$ . The truth of my belief that  $p$  in  $W^*$  is a good example of accidental truth. All of the examples I rely on below are equally unproblematic.

So I do not know that  $p$  in  $W^*$ . But it seems obvious that I know that no dinosaur will appear in the hallway outside of my office in one second; that is, in  $W$  this is true, and it seems obvious that I know it. It seems, therefore, that I also know that  $p$  in  $W$ . Ex hypothesi  $W$  and  $W^*$  share an initial segment up to and including time  $t$ ;<sup>3</sup> furthermore, in both  $W$  and  $W^*$ , at  $t$ , I truly believe that  $p$ . But it seems that in  $W$  I know that  $p$ , and in  $W^*$  I do not. (1), so it seems, is false. This is the first horn of the dilemma.

Suppose that (1) is false. If so, then (for at least some propositions) if I know that  $p$  at time  $t$  in some possible world, then there is another possible world exactly like that one in every way up to  $t$  in which I truly believe, but do not know, that  $p$ . So it could be that I now know that  $p$  although in some other possible world, exactly like this one in every way up to the present time, I truly believe that  $p$ , but do not know that  $p$ . If (1) is false, then whether a true belief is knowledge or not is not wholly a function of what has occurred up to and including the time of belief.<sup>4</sup>

We already know that whether or not a belief is true could depend on what has yet to happen; from this it follows that whether or not some of my current beliefs are knowledge could turn on future happenings; so, one might wonder, what new conclusion is supposed to follow from a rejection of (1)? The answer is that rejecting

(1) has important consequences for our theorizing about knowledge, consequences that do not follow from the fact that my belief's truth could depend on future occurrences. For instance, the denial of (1) appears to be inconsistent with any account of what makes true belief knowledge that both depends exclusively on justification and also understands justification in terms of evidence or reasons for belief. Could one seriously claim that whether my evidence or reasons for belief are any good (whether my belief is, in this sense, justified) could turn on what has yet to happen? After all, if I now have good reasons for my belief that p, then someone in exactly the same epistemic position as I am, now confronted with all the same evidence, now having all the same beliefs, must have equally good reason for believing that p; what happens in the future is irrelevant. (Note that such theories of knowledge are not similarly challenged by the mere fact that whether my belief that p is true could depend on what has yet to happen.)

Likewise, rejecting (1) presents a problem for one who thinks of knowledge as justified true belief, and understands justification in terms of deontology. How could what has yet to happen determine whether or not I am now doing my epistemic duty with respect to believing that p? This would make the demands of deontic justification doubly burdensome: not only am I required to engage in a certain sort of epistemic behavior; but whether my behavior is now appropriate, and thus whether I now properly avoid censure, depends on what has yet to happen!

Similar problems arise for one who claims that what makes true belief knowledge is coherence, but in accepting the denial of (1) adds that coherence must embrace not only the beliefs I now have (and have had), but the beliefs I will have in the future as well. What exactly coherence amounts to will, of course, vary from account to account. But it is surprising that a coherentist of any stripe would claim that, although my beliefs are now, and always have been, maximally coherent, my current true belief that p might fail to be knowledge. Certainly the internalist appeal of coherence theories of knowledge disappears once coherence becomes a function of, among other things, beliefs I do not yet have.

This last remark suggests that problems raised by rejecting (1) are particularly acute for anyone with internalist inclinations.<sup>5</sup> Rejecting (1) entails that what makes my true beliefs knowledge is a function of what has not yet occurred, and no matter which factors are included in what is internal, presumably what has not yet occurred must be excluded.

Of course, it is already well established that the justified true belief analysis of knowledge is inadequate. And it is possible, one might add, that the fabled "fourth condition" that is meant to handle the Gettier problems that plague this analysis, and is presumably externalist in nature, might also handle the problems raised by a denial of (1). So, for instance, maybe S knows that p if and only if S's belief that p is justified (in a manner amenable to internalists) and some further condition is satisfied; and maybe the satisfaction of that further condition takes care of, among other things, just the sort of problems with future beliefs that have been our focus.

So it might be that rejecting (1) is consistent with accounts of knowledge in which justification plays an important role, even if justification is given an internalist analysis.

What then, one might ask, follows from the falsity of (1) other than the widely accepted claim that a fourth condition, presumably externalist in nature, must be added to an internalist's account of knowledge? The answer is that rejecting (1) shows us that among all the other elements that presumably must go into this fourth condition are facts about what has not yet occurred. The significance of this is highlighted by noting that not every paradigmatically external factor will be a candidate for this role. Since presumably causation runs from the past to the future, it seems we cannot avoid the problems that rejecting (1) raises for internalist accounts by relying on the fact that one's belief is caused in the right manner. So even if one already accepts an externalist "fourth condition" to handle Gettier problems, rejecting (1) still has a surprising result: that fourth condition must somehow make room for future events, and might not even be satisfied by something as robustly externalist as proper cause of belief (but see the discussion of causation below).

Note that if this is right, then rejecting (1) should trouble not only the internalist, but many externalists as well. For example, consider the thesis that S knows that p if and only if p is true and S's belief that p is formed by a reliable belief forming mechanism. This, combined with the denial of (1), leads to the further claim that whether or not the mechanism that forms one's belief is now reliable can turn on what has yet to happen. If so, then whether or not I now reliably form beliefs about, for instance, where dinosaurs will or will not be could depend on whether dinosaurs start popping into existence in the future. If they do, then perhaps the relevant belief forming mechanisms are now not reliable (or are not reliable when directed at the future). If, on the other hand, no such miracles occur, then perhaps the mechanisms are now reliable.

This is not a wholly implausible emendation to reliabilism, but neither is it trivial. If one is forced to construe reliability in terms that somehow take future events into consideration, then reliability of a mechanism cannot possibly be understood wholly in terms of its past "track record". So, for instance, the fact that a particular mechanism has formed all and only true beliefs on a thousand separate occasions cannot entail that the mechanism is reliable; whether and how that mechanism will perform in the future, or would perform regarding future-directed beliefs, must also be factored in. Regardless of one's view of the merits of "track record" accounts of reliability, it would be rather surprising if they were ruled out, not by reflection on reliability as such or even on the details of a theory that makes heavy use of reliability, but rather by the fact that, in some possible world, I falsely believe that a dinosaur will not appear in the hallway and disjoin this with "Brown is in Barcelona".

A similar point applies to a theory of knowledge in terms of causation generally. Suppose one thinks that knowledge is true belief caused in some particular way. Combining a theory of this sort with a rejection of (1) entails that whether my belief was caused in a certain way can depend on what has yet to happen. One who thinks, for instance, of causation in terms of past, present, and future regularities will have no objection to this conclusion. Everyone, however, should be surprised if such a substantive claim about the nature of causation - that, roughly, which causal laws are now in play depends (in part) on what has yet to happen - should follow not from considerations about causation as such, but from facts about beliefs I possibly have about dinosaurs or Brown's location. Of course, accommodating a rejection of (1) by making the current causal laws depend (in part) on the future is absolutely unacceptable to one who understands knowledge in terms of true belief appropriately caused, but thinks that which causal laws currently reign is broadly logically independent of what has yet to occur. This would include, for instance, anyone who holds that there is some possible world (however distant) in which L is at one time a law of nature, but that, by some miracle, L is not a law at some later time.

There are many ways in which external factors, factors to which a knower cannot have "direct access", have been invoked in accounts of knowledge. Epistemologies that, for instance, stress the reliable, or generally causal, origin of belief imply that the past has a role in making true belief knowledge. Others have argued that facts about one's environment - barn facades down the road, sheep over the next hill, newspaper reports one has not read - play a role. There are, of course, other examples that we could add. And if (1) is false, then, in addition to the past, the environment, and other familiar external conditions, there is yet another factor to which we lack direct access that must be included in our theory of what makes true belief knowledge: the course of the future. This is a significant consequence, especially when we realize that the claim is not that one or another particular theory of what makes true belief knowledge is forced into embracing an externalism that includes, at least, the future, but that all theories - regardless of the presumed understandings of causation, regardless of whether they are otherwise generally internalist or externalist - must do so.

So the first horn of the dilemma is this: (1) is false; therefore, whether or not S knows that p at t is a function of, in addition to p's truth, what happens in the world after t. If we embrace this horn of the dilemma, then a significant new constraint is placed on how we approach or understand an analysis of what makes true belief knowledge.

## II

Can we avoid the first horn of the dilemma? Only, obviously, by embracing the second; that is, only by holding that (1) is true. This, of course, commits us to the claim that any purported counterexample to (1) is not a real counterexample. How might we discredit the counterexample to (1) proposed above? Recall that the

example made use of the following claims: two worlds,  $W$  and  $W^*$ , share an initial segment up to, and including, time  $t$ ; in  $W$ , the actual world, I believe that  $p$  (that there will be no dinosaur in the hallway in one second or Brown is in Barcelona); in  $W$ , just as we would expect, no dinosaur appears in the hallway one second after  $t$ ; in  $W^*$   $p$  is true, but only because Brown happens, unbeknownst to me, to be in Barcelona. The proposition in question is: (1) If two worlds share an initial segment up to, and including time  $t$ , and if at  $t$   $S$  truly believes that  $p$  in both worlds, then at  $t$   $S$  knows that  $p$  in one world if and only if  $S$  knows that  $p$  in the other. It is obvious that in  $W^*$  I do not know that  $P$ . So the only way to deny that we have a genuine counterexample to (1) is to insist that, since (1) is true, I do not know that  $p$  in  $W$  (the actual world). This might not seem so implausible: one might hold that I do know that there will be no dinosaur in the hallway in one second; but since, one might argue, knowledge is not closed under disjunction introduction, I fail to know (the disjunctive)  $p$ .

But not all purported counterexamples to (1) rely on disjunction introduction. There is a genuine counterexample to (1), and thus (1) is false, if there is a possible world where both:  $S$  knows that  $p$  at  $t$ ; and all that has happened up to, and at,  $t$  does not entail that, at  $t$ ,  $S$ 's belief that  $p$  is not accidentally true. If there is such a possible world, then there is another world that shares an initial segment with the first up to, and including, time  $t$ , but in which, at  $t$ ,  $S$ 's belief that  $p$  is true but accidentally so (and therefore not known). But this, of course, is just what (1) asserts is not possible. Therefore, anyone who wishes to defend (1) must assert that, necessarily, if  $S$  knows that  $p$  at  $t$ , then all that has happened up to, and including,  $t$  entails that, at  $t$ ,  $S$ 's belief that  $p$  is not accidentally true. This is equivalent to the claim that if it is possible, given all that has occurred up to and including time  $t$ , that at  $t$   $S$ 's belief that  $p$  be accidentally true, then, at  $t$ ,  $S$  does not know that  $p$ . In other words: (2) If two worlds share an initial segment up to, and including, time  $t$ , and if in one of the worlds, at  $t$ ,  $S$ 's belief that  $p$  is accidentally true, then, at  $t$ ,  $S$  does not know that  $p$  in either world. So (1) entails (2). But (2) dramatically curtails the extent of our knowledge of the future. Suppose I now believe there will be no dinosaur in the hallway in one second. There is another possible world that shares an initial segment (up to and including the present time) with this world, but in which my belief is accidentally true. For it is possible that by some miracle a dinosaur appears in the room at the end of the hallway in half a second, and then hesitates a full second before walking into the hallway. In that possible world my belief that there will be no dinosaur in the hallway in one second is true, but its truth is a happy coincidence. In that world, I do not know that there will be no dinosaur in the hallway in one second. Therefore, since (1) entails (2), the defender of (1) must conclude that in this world I do not know that there will be no dinosaur in the hallway in one second.

Further examples of what (2) entails that I cannot know: that my wife will not be abducted by aliens this afternoon (it is broadly logically possible that aliens come into existence this afternoon and attempt the abduction, but, against all odds, they are thwarted); that my body will be within the Milky Way Galaxy at lunch time (it is

possible that somehow my body is soon hurtled into space, and at lunch time is just on the fringes of the galaxy, heading out); that I will not be a brain in a vat by midnight (it is possible that my brain be placed in a vat at 11:50 this evening, and the vat break at 11:55); or that my best friend will not murder me tonight in my sleep (it is possible that she try, but that she succeed in only wounding me). Further, I will obviously fail to know any proposition that is such that knowing it entails that I know one of the propositions ruled out by (2); for instance, if it would take me a second to step to the door and look into the hallway, then, presumably, I don't know that if I were to look in the hallway, I wouldn't see a dinosaur there.

Furthermore, if (1) (and therefore (2)) is true, induction cannot be a source of knowledge. Among the things we cannot know about the future is that it will be like the past in certain important ways. Presumably it is possible that the future will be like the past in these ways, but because of the whim of an evil demon (who could possibly come into existence at any time after the inductive belief is formed). In such a case we could hardly be said to know that the future will thus resemble the past.<sup>6</sup>

In arguing that (1) leads to (2) (and in my defense of the first horn of the dilemma), I have presupposed that future contingent propositions - propositions like there will be no dinosaur in the hallway - have truth values. One might deny this. However, the claim that all future contingents lack truth values entails that none of them is true, and therefore, since a belief is known only if it is true, entails that none of them is known. To deny that future contingents have truth values is to embrace the second horn of the dilemma; we cannot know many things about the future that we intuitively think we do know.<sup>7</sup>

In conclusion, any theory of knowledge faces the following dilemma: either (1) is false, in which case whether or not what one truly believes now is knowledge is a function of, among other things, what has yet to happen; or (1) is true, in which case one can have surprisingly little knowledge of the future.<sup>8</sup>

## BIBLIOGRAPHY

Chisholm, Roderick (1975), *The Problem of the Criterion* (Milwaukee: Marquette University Press).

Dretske, Fred (1971), "Conclusive Reasons," *Australasian Journal of Philosophy*, vol. 49 (1971):1-22.

Nozick, Robert (1981), *Philosophical Explanations* (Oxford: Clarendon Press).

## NOTES

(1.) If God exists, and if God's beliefs about the future are "hard facts", then we need to exclude God's future-directed beliefs (and intentions) from what is shared in an initial segment.

(2.) (1) is, of course, implicitly universally quantified: it is meant to hold for any world, person, time and belief. Because the truth of a proposition about the future is a soft fact - indeed, a paradigm of a soft fact - S's truly believing that p must be added to, and is not included in, the claim that the two worlds share an initial segment up to, and including, time t. Although S's believing that p at t is a hard fact, and thus shared by the worlds in virtue of sharing an initial segment which includes t, the truth of p, if p is about what will happen after t, is not.

(3.) They differ, of course, in that before t in W I have the true belief that there will be no dinosaur in the hallway, and in W\* this belief is false. But my belief about the dinosaur is a belief about what the world will be like after t - its truth, therefore, is a paradigmatic soft fact and not included in the shared initial segment up to, and including, t (my believing that there will be no dinosaur is a hard fact and included in the shared segment).

(4.) Merely counterfactual facts about a world before, or at, time t will not necessarily be included in its initial segment up to, and including, time t. Imagine that S believes at t some future-directed proposition p (such as: it will rain tomorrow). The counterfactual if p were false, S would not believe that p, combined with the hard fact that, at t, S believes that p, entails that p is true. Such counterfactuals (when combined with what is clearly a hard fact) entail what is clearly not part of the initial segment up to and including time t; therefore, these sorts of counterfactuals regarding beliefs about the world after t are themselves not included in the initial segment up to and including time t. Counterfactual analyses of knowledge are endorsed by Fred Dretske in "Conclusive Reasons" and Robert Nozick in *Philosophical Explanations* (p. 172).

(5.) I make use of a familiar, but rather vague, distinction between internalist and externalist theories of knowledge. Roughly, the distinction is that according to internalists, whatever plays the central role in making true belief knowledge is something to which each person has, in his or her own case, an immediate and perhaps infallible access. The externalist denies this. As will be acknowledged in the discussion below, the disagreement between the two camps is really a matter of degree; most internalists agree that there is some externalist component to what epistemizes true belief.

(6.) None of this is to deny that, even if (1) is true, we can have justified beliefs about the future, or to deny that it is rational to act on beliefs about the future. But if (1) is true, then we get the quite surprising result that we cannot know facts about the future such as that there will be no dinosaur in the hallway in one second. This should be especially troubling to anyone who favors a "particularist" (as opposed to a "methodist") approach to epistemology. According to the particularist, in formulating our theory of knowledge, we start out with certain paradigm cases of what we know, and try to work out a theory that saves these paradigms. I would have thought, however, that claims like "there will be no dinosaur in the hallway in



one second" or "if I turn around, I won't see a dinosaur" are paradigmatic examples of knowledge. Roderick Chisholm introduces particularism and methodism in *The Problem of the Criterion* (p. 15).

(7.) There are some putative cases of knowledge about the future that (2) does not seem to rule out. Some examples: one's belief that  $2+2$  will equal 4 in the year 2001; one's belief that tomorrow will be such that it was raining today; one's beliefs about the future if one essentially has true beliefs about the future; and one's beliefs about the future if one comes to hold those beliefs because one was told by an essentially omniscient, essentially honest being that they were true. (8.) Thanks to Marian David, David Lewis, Eugene Mills, Mark C. Murphy, Alvin Plantinga, Philip Quinn, Leopold Stubenberg, Dean Zimmerman, and an anonymous referee for helpful comments.