

Equivalent Worlds and Knowledge

Kristian D'Amato

November 2011

Three Aberrant Cases

One disturbing but telling demonstration of how definitions of knowledge can go wrong is due to Hawthorne.[1] Although properly intended as an attack on Dretske's theory of conclusive reasons, it also works against Nozick's conditional theory. Hawthorne provides a simple example: suppose, he declaims, that salmon induces hallucinations if eaten in great amounts. Suppose that, to be more precise, were you to eat it in amounts greater than 14 pounds at one go (never mind how *that* is humanly possible), it could cause you to believe that you've only had a few tidbits of the savoury fish. Then, on Dretske's account, if **P** is the proposition **I ate less than 1 pound of salmon**, I could know **P**. On the other hand, let **Q** be the proposition **I ate less than 14 pounds of salmon**. **Q** is not knowable, because were **Q** to be false, I might still experience the hallucination that I've eaten less than a pound of salmon, and therefore I do not have conclusive reasons for **Q**.

A similar claim could be made within Nozick's theory: since my belief in **Q** is not negated by **Q**'s falsity, the belief does not satisfy one of the necessary conditions of knowledge according to that account. The strangeness of this result is not alleviated by the seemingly straightforward inference that if I ate less than a pound of salmon, then I must have surely eaten less than fourteen times as much! In everyday life, unfortunately, you would be hard-pressed to find someone who places doubt in that implication, but knowledge is not closed under entailment in either Nozick's or Dretske's theory. Incidentally, closure is a subject that we shall return to later; at this juncture I simply want to point out the conflict with common sense, resulting, I believe, not necessarily because of something wrong in the respective definitions of knowledge, as the example purportedly shows, but from a subtly different metaphysical presupposition with regard to what beliefs in propositions **P** and **Q** are taken to mean.

Another aberrant case that might throw some light onto the problem, albeit in a very different manner, is the following. Imagine a person **A** that has been hooked up to a hallucination-inducing device under the control of a scientist. The latter tweaks the machine so that when an apple is placed in a certain spot in front of **A**, **A** experiences a hallucination of an apple in the exact same location as the real apple. So if there *really* is an apple in that particular spot, **A** will see an apple. If not, he won't. If **P** is the proposition **There is an apple before me**, then it satisfies both subjunctive conditionals because it tracks the truth perfectly, so **A** can know **P**

according to Nozick. But would we be so quick as to give it the status of knowledge, and if we do, what accounts for that undeniable hesitation to do so?

A third and final case is a sceptical move owed to Bertrand Russell:

There is no logical impossibility in the hypothesis that the world sprang into being five minutes ago, exactly as it then was, with a population that “remembered” a wholly unreal past. There is no logically necessary connection between events at different times; therefore nothing that is happening now or will happen in the future can disprove the hypothesis that the world began five minutes ago.[2]

Not only is a five-minute-old world not necessarily logically impossible, but neither is a six-minute-old world. Nor, for that matter, is any x -minute-old world, where x is a member of the real numbers that are greater than zero. There are, of course, an infinite number of such worlds, and an infinite number of possible beliefs of the form $P_x = \text{The world was not created } x \text{ minutes ago}$ that correspond to each of these worlds, despite the fact that having a serious doubt in any of them is not a healthy sign. Yet they pose an interesting philosophical problem; moreover, no member of this family of related beliefs is knowable in the conditional theory. For, the standard response goes, were P_x to be false, I would not believe it, because the experience would be entirely identical to the one I currently have (i.e. with the world apparently billions of years old).

Equivalent Worlds

Common to these three examples, and perhaps crucial to their recognition as anomalous cases, is the notion of identical experiences. It may be the case that P is true, for which I have supporting evidence E , but deviant case Q is also corroborated by the same evidence E —so which am I to believe? It is thus unsurprising to conclude that identical sensory experiences expose a perplexing issue in epistemology, and it seems to me that we often run the risk of taking for granted certain metaphysical beliefs in attempting to exorcise these most pertinacious demons of scepticism. The situation is made even more frustrating because the formidable argument from error is fueled by such cases, sending epistemologists into a fit of desperate measures in order to dodge the sceptical bullet.

I believe, however, that an investigation of knowledge, or rather, a rethinking of our views on what knowledge achieves, may be sufficient to dispel the argument from error. In view of this end, let us for the sake of clear designation refer to possible worlds that generate identical experiences as *equivalent worlds*. A universe that is 13.6 billion years old and a universe that is only five minutes past its genesis, complete with faux memories and fossils, are then equivalent worlds. So are the case where I've only eaten half a pound of salmon and the case where it was actually fifteen, or the case where I'm experiencing a real apple as opposed to a hallucination of one in the same spot.

A second observation about these cases is that there is something tenuously different between the implications of a claim like **I know that I ate jacket potatoes**

yesterday evening, for example, and the claim **I know that the universe was not created five minutes ago**, and I do not solely mean the obvious discrepancy in terms of the vastness of each respective claim, but one in their *quality*. The same could be said of the claim **I know I am sitting down typing this essay at 12:57 in the morning** and the claim **I know I am not experiencing a dream to that effect**. In this example the distinction is clearer; it seems to be one in the fundamental nature of the propositions—we would only be scratching the surface by claiming that the former confirms observations from immediate experience, while the latter describes a more profound form of knowledge.

In light of the previous two observations and the problems that give rise to them, it would perhaps be an interesting manoeuvre to think of knowledge not as an affirmation of any particular world, but as an affirmation of a *set* of equivalent worlds. Intuitively, it is straightforward to grasp the idea that the knowledge claim **I know that there is a computer monitor before me** expresses my position on the subject of possible worlds as follows: the *actual* world is one of a multitude of possible worlds in which there is a computer monitor before me. My knowledge claim leaves much space for undefined (or ill-defined, one could say) possibilities; it does not confirm or deny anything about what lies behind my computer monitor, for instance. Nor does it claim that the monitor does not mischievously blink out of existence when nobody is paying attention.

This knowledge, if it amounts to that, places little restriction on the universal set of possible worlds—to borrow some mathematical jargon, we could equally say that the resulting equivalence class in which my computer monitor proposition is true is relatively large, or coarse-grained; it contains all worlds in which I could truthfully hold my belief, which as far as beliefs go does not add up to much. If, on the other hand, I found myself in a position to affirm all there is to affirm on the basis of evidence about the world,¹ a quasi-omniscient position in some ways akin to that of Laplace's demon, but one that does not restrict its intellect to matters of purely physical phenomena, the equivalence class consisting of all possible worlds consistent with what I know would be much smaller, or fine-grained.

Still, there is room for play, in that my quasi-omniscient powers of observation could not differentiate between individual elements in the set of equivalent worlds, however small it is made to be by my extensive awareness. Like the set of real numbers, in which there are an infinite collection of numbers between any two distinct elements, so does this putative set of equivalent worlds admit an infinite number of possibilities consistent with what I experience, at least at face value.² Thus, I can know all there is to know and yet be unable to distinguish between the hapless world that was created a few minutes ago and that which traces its lengthy story back through eons of time to the Big Bang.

The strength of this suggestion is that it lays open a route to understanding what

¹That is, by reasoning from empirical observation, and forming justified abstractions of all that could be available to a healthy rational mind.

²Of course, not all of these possibilities will be equally sensible, but the reasons for that might not be immediately available to experience; rather, they would be justified conclusions reached by the rational mind, that from amongst the continuum of often ludicrous possibilities picks out some and not others. Unfortunately, a discussion of what justifies such positions is beyond the scope of this essay.

I mean by claiming to know **I ate jacket potatoes yesterday evening**: there exists a dichotomy of the possible worlds, a partition that splits them into those consistent with (that is, experientially equivalent to) my statement, and those that are not, and the evidence that I have at hand affirms the actual world as being one that lies within the former group. It does not make a claim as to *which* of those worlds is the actual world; my understanding leaves intact the possibility that yesterday's meal is only a memory injected into my mind by a devious deity at the moment that he created the world five minutes ago, or that I am fooled into thinking it because of my less-than-perfect memory. These possibilities are embodied by elements in the equivalence class of worlds consistent with the truth of my proposition.

Alright, so can I know that the universe was in fact created in an event that occurred billions of years ago, as generally asserted by modern physicists? If when I state this proposition I allow for any and all worlds borne by the empirical evidence that scientific endeavour has made available, then the answer is yes, I can know it, because I am not interpreting it as a metaphysical claim. If, however, I do arrogate the metaphysical assertion that on the basis of Big Bang evidence alone, the universe was indeed created billions of years ago, and reject the possibility of a metaphysical blunder, then I do not, *a prima facie*, know that claim. For then I would need a theory of metaphysics consistent not only with Big Bang evidence, but with the entire motley collection of empirical evidence. In other words, the evidence can be seen to confirm not one scientific theory, but a whole family of them that are compatible with the facts. Now, even if there were only one scientific theory contending for a description of the universe, I could still think of an infinite number of metaphysical claims that, although mutually exclusive with respect to each other and the scientific theory, would give rise to the same collection of evidence as we outlined above. To select from amongst those claims, I would need some sort of guiding principle whose validity rests not simply on the body of evidence that pertains to the creation of the universe, but on the entire corpus of knowledge.³

This may seem like it plays directly into the hands of the sceptic, who'd have a field day with our apparent incapacity to choose between a promising scientific theory and a fraudulent bit of metaphysical humbug, but it does not. For knowledge is useful in so far as it partitions the universal set of possibilities; i.e. when we ponder the truth or otherwise of a proposition, we are usually concerned about the entire set of equivalences that reflect on that proposition, not any particular world. So when I worry, say, about neutrinos breaking the speed of light, I am worrying about the whole set of possible worlds in which neutrinos travel faster than the speed of light—including worlds where apes can talk, worlds where people think it absolutely normal to hop their way to work, etc. And that includes worlds where there is an evil demon that makes it appear as if neutrinos travel faster than light. Why does our concern usually extend to the entire equivalence class? It has to do with how we assign meaning to terms and subjects of systematic study. We *define* neutrinos to be certain things; no metaphysical truth unveils itself to the physicist when she observes the neutrinos' effect in passing. It's just that certain observations in certain

³Occam's Razor, to hazard a guess, might be such a principle; again, a discussion of such principles is beyond the scope of the present essay.

conditions are assigned to certain objects that we call certain things, and this is how we make sense of the world.

On this view, then, the sceptic is misguided in that he views knowledge as an attempt to fix our world onto a single indivisible possibility. When this attempt fails, as it is apt to do, given the leeway that our evidence tolerates, he throws down the gauntlet and repudiates our claims to knowledge. In fact the sceptic is fundamentally mistaken: knowledge is mainly a device useful in making the world intelligible in terms of groups of similar possibilities and their interrelationships, not one that clinches the universe to an individual immutable set of truths.

The Closure Principle in the Conditional Theory of Knowledge

One appealing factor of the above view is that it explains some of the disputes that revolve around the closure principle under known entailment: why Nozick's theory holds that it is false, for instance. It seems to me that most purported demonstrations of the principle's failure are guilty of the crime that I alluded to in my second observation before, an inadvertent though unrestrained mixture of terms that implicitly involve appearance and reality in such a way as to render them incompatible. Nozick's employment of counterexamples is a case in point.

Let us use Descartes' dream argument by way of explication. Nozick argues that I can know that I am sitting reading because such a belief easily satisfies the two subjunctive conditionals that are necessary for knowledge. This implies that I am not dreaming, an implication that I can know. However, notwithstanding my knowledge of the latter, I cannot know the implied proposition, namely that I am not dreaming, because if I *were* dreaming, I would not necessarily believe that I am! [3]⁴

Why does the closure principle fail in this scenario? On my view, it does because the straightforward claim from appearance $\mathbf{P}_a = \mathbf{I\ am\ sitting\ reading}$ is confused with a claim that is expressed in the same words but that means a substantially different thing, specifically the metaphysical claim that I am sitting reading, ruling out all experientially equivalent worlds. Let us call the second belief \mathbf{P}_r .⁵ Unfortunately, given my evidence, I can only truthfully claim to know \mathbf{P}_a , since I could still have believed \mathbf{P}_r had I, say, been dreaming that I am reading; thus, the conditional $\sim\mathbf{P}_r \rightarrow \sim\mathbf{BaP}_r$, where \mathbf{a} is myself, is not satisfied. I can *only* know a proposition if the claim embraces the whole of the equivalence class that is consistent with that proposition, since I have no way of telling apart individual worlds in that class. Then \mathbf{P}_a is knowable, whereas \mathbf{P}_r is not.

Now, the implication $\mathbf{P}_r \rightarrow \mathbf{Q}_r$, where \mathbf{Q}_r is the metaphysical claim **I am not dreaming** (that is, I am not dreaming, and it does not simply *appear* to me as if I'm not dreaming, as it sometimes does in lucid dreams) arguably holds and is knowable, since the antecedent logically precludes the possibility that I am dreaming. The crux lies here, however: if we replace \mathbf{P}_r with \mathbf{P}_a , the implication does not hold anymore, since I could be happily dreaming of sitting at my desk reading, giving $\mathbf{P}_a \& \sim\mathbf{Q}_r$! The upshot is that we have reduced the example to the following sym-

⁴This last is reminiscent of Hawthorne's salmon example.

⁵The subscripts 'a' and 'r' are being used in allusion to appearance and reality respectively.

bolic representation: $[\mathbf{KaP_a} \ \& \ \mathbf{Ka(P_a \rightarrow Q_r)}] \rightarrow \mathbf{KaQ_r}$, which, as we have seen, can never have a true antecedent on account of the implication not being knowable. A final attempt at slipping by would be the closure principle expressed as $[\mathbf{KaP_r} \ \& \ \mathbf{Ka(P_r \rightarrow Q_r)}] \rightarrow \mathbf{KaQ_r}$, but this manifestly fails on the grounds that $\mathbf{P_r}$ is then unknowable. The closure principle emerges intact.

References

- [1] Hawthorne, J. (2005). The Case for Closure. In M. Steup & E. Sosa (Eds.), *Contemporary Debates in Epistemology*. Oxford: Blackwell Publishing.
- [2] Russell, B. (2007). *The Analysis of Mind* (Rev. ed.). London: Routledge.
- [3] Dancy, J. (1985). *An Introduction to Epistemology*. Oxford: Blackwell Publishing.