



John Hawthorne

Religious Knowledge¹

John Hawthorne

Let me distinguish two sorts of epistemological projects that might be undertaken by someone with religious commitments. The first – roughly an evangelical project – is that of trying to get other people who do not currently share those commitments to come around to one's point of view. I am not interested in that project here. A second project is rather more solipsistic. It is that of resisting internal pressures to admit that even by one's own lights, it is overwhelmingly unlikely that one's religious commitments are rational and overwhelmingly unlikely that one's religious commitments constitute pieces of knowledge. The project here is not that of convincing others – it is that of keeping the epistemological wolves at bay.

The second project is an altogether worthy one. There, is after all, something very uncomfortable about being in the situation where one maintains an unqualified commitment to some proposition P but is also committed to either the proposition that one doesn't know P or to the proposition that one is irrational to believe P. Speeches of the form 'P and I don't know P' and 'P and I am irrational in believing P' sound none too healthy, whether made to others or to oneself.

Now perhaps there are some people who are so unreflective that they feel no internal pressure whatsoever towards denying that their religious beliefs are rational/pieces of knowledge. These wide eyed individuals claim to know this or that but have thought very little about what it takes to know, and about what sorts of situations provide hostile environments for knowledge. They will feel no need to bring the second project to a satisfactory conclusion. This paper is not directed at them.

I shall be looking briefly at two strategies by which religious believers might attempt to ward off pressures towards a negative epistemological self-assessment. Both are fairly well known. The first is a 'fine-tuning' strategy, wherein certain empirical claims about fine-tuning are used to justify claims to the effect that high credence in theism is rational. The second is a 'trust' strategy, which relies on the claim that a belief that P that flows from trust in an authority who knows that P itself constitutes knowledge, whether or not there are proximate misleading authorities, and whether or not the relevant authority also communicates falsehoods.

(1) *Fine-tuning*

Prima facie, it seems pretty clear that whereas the best scientific theories of, say, the fourteenth century, provided a very friendly environment for religious belief, current scientific theory provides a rather unfriendly environment. Six or seven centuries ago there were all sorts of features of physics and cosmology that were most encouraging to the religious believer. Aristotelian philosophy of motion told us that bodies ran out of steam unless they were given fresh impetus. This led naturally to the thought that the universe as a whole would be running out of steam unless it was continually being given impetus from the outside. (Note that this ‘first cause’ argument does not require that the universe had a beginning. An eternal universe would hardly alleviate the need for motion to be inputted from the outside as a means to keeping the cosmological wheels in motion.) Meanwhile, Ptolemaic astronomy told us that we were literally at the centre of things, a fact that would seem strange and inexplicable from the perspective of a barren atoms-in-the void minimalism. And Aristotelian metaphysics of nature told us that the development of things was to be explained in terms of their principle of flourishing – their substantial form – and could not be explained by their material make up. This perspective reinforced the priority of teleological over mechanical explanation. These and many of the other sources of encouragement for religious belief that were available back then have by now evaporated. Current scientific theory seems to provide no encouragement whatsoever to the religious believer.

Now one reaction that the religious believer might have to this is a don’t care attitude – we shall look at one version of this in the next section. Another reaction is that of trying to find subtle grounds for encouragement in current scientific theory, claiming that, under close scrutiny, science continues to provide a very friendly environment for religious belief. The best known version of this response is the fine-tuning strategy to which I now turn.²

The strategy I have in mind relies on an empirical springboard which I shall not question. This key empirical claim is that for an important range of fundamental physical constants c (Planck’s constant, for example), something like the following is true:

Only an extremely restricted range of values for c permit the existence of living conscious beings.

(The key bit of empirical reasoning demonstrates that for any of a myriad range of settings for this or that constant, those settings would make it impossible for large stable physical objects to emerge out of the cosmological soup.)

The line of thought then proceeds by claiming that the fact that the actual

world's constants are set at c provides a compelling reason for preferring the hypothesis that the creation of the universe was manipulated by a being who was interested in the emergence of living conscious beings over the hypothesis that the universe has no manipulator.

The line of thought can be made vivid by the following analogy. Suppose I am presented with two buckets of pills. One bucket is the benevolent bucket – it contains pills that put me to sleep for a week after which I wake up refreshed. Another bucket is the risky bucket. It contains a few pills of the sort contained in the benevolent bucket, but the rest of the pills make me unconscious for ever. Suppose I have no idea which bucket is which. I take a pill from a bucket. At that point I have no view as to whether I have taken a pill from the benevolent bucket or not. But suppose at some point I find myself waking up (memory intact). At that point I clearly have excellent reason to believe that I have taken a pill from the benevolent bucket. Let a pill represent a setting of the universe's constants, where a waking pill represents a setting for the constants that allows for conscious beings and an unconsciousness pill represents one that does not. The benevolent bucket represents the hypothesis of a consciousness-desiring manipulator, since such a manipulator would always ensure conscious-friendly constants, while the risky bucket represents the hypothesis of a blind universe. Meanwhile, the fact of waking up represents our living conscious existence. Just as in the scenario described I give high credence on waking up to the claim that I have taken a pill from the benevolent bucket, so I am supposed to give high credence to theism on taking cognizance of the relationship of constants to consciousness in tandem with my own conscious existence.

One kind of reaction to this kind of strategy is to contest the claim that the relevant empirical facts in combination with my own conscious existence provides evidence for theism. I think this is the wrong reaction. To make this vivid consider a much more silly example. Consider the hypothesis that the universe has been manipulated by a being whose main ambition is to produce a universe with large orange vegetable. Let us ask whether the existence of pumpkins is evidence for the existence of such a being. Well, let us ask what credences we ought to have had for the relevant hypothesis prior and posterior to noticing pumpkins. On standard ways of thinking about how probabilities are updated,³ the probability assigned to the hypothesis would go up upon noticing pumpkins. To simplify, suppose we are 50/50 whether there are large orange vegetables in the world and think it one in two hundred billion that the world is manipulated by a being obsessed with large orange vegetables. When we discover there are pumpkins, then on standard ways of thinking about updating (and supposing there is no other relevant new evidence) we ought to be one in a hundred billion that there is such a being.

Think of someone's pattern of credences as a pie, with the size of credence

in some proposition corresponding to the size of some associated piece (relative to the whole pie), where the piece representing credence in *P* is part of the piece representing credence in *Q* if and only if *P* entails *Q*. New evidence consists in knocking out certain pieces. Standard updating is tantamount to expanding the pieces that are left into a whole pie in a way that preserves the ratios between them. When we learn that there are large orange vegetables, we destroy half of the pie. When the remaining half is expanded to form a whole, the piece that used to constitute one two hundred billionth of the pie now constitutes one hundred billionth of it.

(If one wants a bucket analogy, imagine that there are billions of buckets, one of which is full of waking pills, the rest a mixture of pills, half of which lead to waking, half to permanent unconsciousness. One takes a pill, not knowing for certain which kind of bucket one is taking a pill from. When one wakes up, that provides a bit of evidence that one has taken a pill from the bucket full of waking pills, but one's credence in that hypothesis remains extremely low.)

The thought experiment encourages one to think that the existence of large orange vegetables is evidence for our bizarre supernatural hypothesis. But it hardly encourages one to think it is rational to assign high credence to that hypothesis. This points to the main shortcoming of the fine-tuning strategy. For that strategy focuses on the claim that consciousness plus fine-tuning provides evidence for theism but ignores the crucial issue, that of the appropriate prior probabilities. To get at this issue we might imagine a non-worldly being who was about to be thrust into the universe knowing nothing about it, though who already knows enough about the dependence of physical layout on physical law to know about the various dependences of this or that structure on this or that setting of constants. What kinds of credences are appropriate for that being with regard to the hypothesis of a consciousness-obsessed manipulator? Certainly, the piece of pie corresponding to that credence will be entirely contained within the piece of pie corresponding to the credence in consciousness friendly constants. But the key issue is the appropriate ratio between the former piece and the latter piece. Suppose the proportion is rationally mandated to be very low. Then, when the being learns about consciousness friendly constants, its credence in a consciousness-obsessed manipulator will go up but will remain very low. Presumably when that being has absorbed the facts about the presence of consciousness in the world it will be apprised of all the relevant information fine-tuning enthusiasts produce. So if its initial credences were more or less mandatory and its final assessment of the consciousness friendly manipulator hypothesis is low then unless the fine-tuning enthusiast has additional evidence, his or her final assessment ought to be low as well.

Of course none of this requires the fine-tuning enthusiast to be destabilized. If he or she honestly thinks that the being I described should, prior to entering

the physical universe, proportion her pie in a way that allows most of the piece corresponding to consciousness friendly constants to be taken up by the piece corresponding to a consciousness friendly manipulator, then she will think rational high credence in such a manipulator explicable by the facts about rational initial outlooks in combination with the facts of fine-tuning.⁴ The trouble is that in my experience there are not many reflective individuals – including reflective religious individuals – who honestly think that it is appropriate for the being to do this unless he or she had additional evidence ready to hand. These individuals will not bring themselves epistemological peace by reflecting on the facts of fine-tuning.

(2) *Trust*

In some circles, there is a use of the term “faith” wherein it is contrasted with knowledge. But as it was used in scholastic circles, the term implied no such contrast. As a pretty good first pass, the insight driving doctrines about faith was that knowledge flowed to individuals who had faith in knowledgeable authorities. To what extent can the religious intellectual use that insight to ward off negative epistemological self-assessment? I’d like to divide our topic into two subtopics. First, let us ask whether the following principles are true

Transfer

If x knows that y asserts P and x comes to believe P by trusting y with respect to P, and y knows P, then x comes to know P.

Maintenance

If x knows P by trusting y and continues to believe P on that basis, then, whatever else happens, x continues to know P.

There are various cases which some people no doubt will take as fairly decisive counterexamples to these principles. I am yet to be convinced that any such cases are in fact decisive. One style of case is one where a knowing subject tells one the truth but there are deceivers lurking nearby who one wouldn’t recognize as such. One might think that this is analogous to seeing a real barn in an environment filled with barn facades – where received wisdom has it that one does not know that one is seeing a barn.⁵ This kind of case doesn’t move me much. Suppose I ask someone the time. The person knows the time and tells me. Suppose there are other people in the room who would lie to me. Is it really so clear here that

I do not know the time? I am lucky in a sense, admittedly. But the question is whether it is the kind of luck that is incompatible with knowledge, or instead merely a case where I am lucky enough to get the opportunity to acquire knowledge. One can't adjudicate between these diagnoses simply by throwing around the term 'luck'.

Another kind of case is one where *x* knows *P*, doesn't know *Q*, tells *y* both *P* and *Q* and *y* comes to believe *P* and *Q*. Here we are invited to conclude that *y* knows neither *P* nor *Q*. This case is also less than convincing. Suppose someone reads the sports results in a paper and believes all that he reads. Suppose there is a misprint regarding the Manchester United result. Does that mean that the person doesn't come to know the Liverpool result (suppose that the latter is not misprinted)? It is far from clear. Another kind of case is one where one is told by a knowing subject that *P*, and trusts that subject even though prior to being told *P* thinks it probable that if that subject says *P* that subject is lying. At least when described in this schematic way I am unconvinced by this 'counterexample' as well. Aren't there cases where we would naturally say that someone came to their senses in trusting someone even though they previously were profoundly distrustful of them? In such a case we do not think that their previous profound distrust stopped them knowing when they did trust. (And of course once they did come to know *P* on that basis then it was no longer likely on their evidence that the person was deceptive about *P*). Another kind of case is one where I come to know *P* by testimony but then get misleading evidence that the testimony is faulty, the idea being that this automatically destroys knowledge. Here again matters are not clear. Suppose a parent knows *P*, tells a child *P* and then all sorts of people tell the child that the parents have messed up, but the child sticks to his guns in believing *P*. Is it really so clear that the child stops knowing *P*?

We seem to have a situation of mixed intuitions about cases on the one hand and a pair of fairly elegant principles that seem to capture part of the important architecture of knowledge on the other. In such cases we should be wary about complicating the principle in response to scattered intuitions about cases. No doubt many philosophers will think that I am forgetting all the epistemological work on defeaters, work driven by the idea that beliefs that I acquire may *defeat* a particular belief's claim to knowledge even though it was previously a bit of knowledge. But as far as I can see, much of that work is of such tragically poor quality that it is not at all clear what can be learnt from it. (Certainly the fact that I believe something incompatible with *P* does not stop my belief that *P* being knowledge. I may, for example, retain the incompatible belief only because I have not done internal housecleaning yet.) Despite the protestations of such philosophers, then, I will proceed on the assumption that Transfer and Maintenance are true. Let us now turn to the question of how self-assessment might proceed in that light.

The following analogy will suffice for conveying the main points that I wish to make. Suppose we live in a post-apocalyptic era and the human race is dying out. One group of us decides to leave a library full of important information about our species and its history. Another devious group decides to leave a library packed with false information. We die out and centuries later Martians come and settle on earth. One group settles in the vicinity of the false library. Another group settles in the vicinity of the true library. Each trusts the books that they read and forms and maintains beliefs on that basis. Assuming Trust and Maintenance are correct, the group that reads the true library comes to know its contents. Relatedly, they come to know that the books that they are reading are true. After all, suppose they read a book that says P, recognize that it says P and come to know P by trusting the book. They know that P and that the book says P, so insofar as they do not know this already can come to know that the book truly says P by a simple deduction (I assume here the plausible closure principle that deductions from known premises produce knowledge.) Do they know that they know propositions learnt from the book? What I have said doesn't settle that. But we can well imagine that the people who wrote the book knew the truth of the Transfer and Maintenance principles and so, for various propositions that they knew that they knew, coupled the claim that P with the claim 'If a reader of this book comes to believe P by trusting this book, then that person thereby comes to know P' (call this conditional C). The Martians reading the true book read P, come to believe P and come to know that by trusting the book they came to believe P. They also come to know C by trusting the book. This all puts them in a position, for very many values of P, not only to know P (by reading P in the book), but to know that they know P (by putting together the conditional in the book with other things that they know).

Suppose the Martians who had read the true library met up with the Martians who had read the false library and noticed the discrepancies. Let us imagine each group divided into two kinds of psychological profiles. There were those who got nervous when the discrepancy showed up and became agnostic about both libraries. And there were those who stuck to their guns. Assuming Transfer and Maintenance, the Martians who read the true library and stuck to their guns continue to know its contents and, for very many values of P, continue to know that they know its contents. Would those Martians be able to know that the Martians who were sticking to their guns from the other library were making a mistake? Assuming Transfer and Maintenance the answer is obviously 'yes'. For various values of P, they would know P, know the other Martians were claiming that not-P and then deduce that the Martians were making a mistake. Now of course the Martians who stuck to their guns with the true library would likely be utterly unsuccessful in any evangelical project with regard to those who stuck to their guns with the false library. But we can imagine that this does not shake

them. They realize that there are lots of cases where one has the information that *P*, but there is no way of sharing that information with someone because he or she is not trusting. (Suppose I remember that Janet came into the room first and you misremember it was John. If you don't trust me and stick to your guns I can't help you epistemologically, but that won't necessarily trouble me.)

Will the Martians who stick to their guns be in a position to recognize any flaws in themselves? What I have said so far might suggest that the Martians who read the good library are not flawed in any way. But I do not at all mean to imply that. A Martian who trusted the good library would recognize that he was psychologically so constituted that if he had first read a coherent but false library he would have trusted that library instead.⁶ He will thus recognize that he is someone with a psychological constitution that could very easily have disastrous consequences. Now this recognition need not, as a causal matter, induce any kind of instability, or loss of faith. After all the stick-to-their guns Martians might be so constituted as to be utterly unmoved in their convictions by a recognition that those convictions were produced by mechanisms that (individuated by their internal operations) could very easily have lead to disaster.

Suppose we were to ask stick-to-your guns Martians: "Doesn't the fact that those mechanisms could so easily have led to disaster show that your beliefs are not actually knowledge?" We might imagine them replying that in the sense relevant to knowledge, a belief formed by trust in a knowing authority is produced by a method different from a belief formed by trust in a false authority, and the fact that a different method could easily have generated a false belief does not epistemologically indict the relevant actual beliefs. They might point out, by way of analogy, that even if dreaming consisted in visual processing that internally duplicated veridical processing and even if it produced real beliefs, the fact of dreaming would not prevent beings from knowing their surroundings when they were awake.

What of the facts about likelihood? Suppose we were to say to a dogmatic Martian: "Wouldn't you concede that it is at least quite unlikely that your internal mechanism is delivering knowledge that *P* (for some disputed *P*)?" We could well imagine the Martian being careful to distinguish questions of objective and epistemic probability. "When I was born it was perhaps objectively unlikely that I would come by the knowledge that I have. But that doesn't trouble me any more than that it was objectively unlikely when I was born that I would be standing on this spot right now (assuming that the world is chancy). Turning to epistemic probability. Is it epistemically probable that I know *P*? In general, if someone knows *Q*, then *Q* is epistemically highly probable for them. But assuming that I know that I know *P*, it is epistemically highly probable that I know *P*. Since I am unwilling to deny that I know that I know *P*, I am unwilling to concede it is epistemically improbable that I know *P*. Of course my opponents will see things

differently. But I am not interested in evangelism.”

Now it seems to me that the stick-to-their guns Martians who trust the good library will achieve a kind of inner epistemological stability that keeps their knowledge in place, albeit one that will be alien to those so constituted as to be made nervous. But there will also be something quite sinister about the mechanism by which they accomplish this, something that can be recognized even by those who dogmatically trust in the good library. For such Martians will recognize that the very mechanisms that deliver them with stable knowledge could also easily generate benighted belief systems, perhaps even ones that induce Martians to destroy each other and themselves, generating ignorant Martians impervious to argument and criticism on account of internal facsimiles of the mechanisms that provide the original Martians with stable knowledge of the world. These mechanisms will do this not at distant possible worlds, but at the world in which they live. If Transfer and Maintenance are correct, then we can recognize that a mechanism can be both deeply flawed and dangerous in this way and also be a conduit to knowledge. The Martians who have read the good library might make this particularly vivid to themselves in the following way. They could imagine themselves given a choice between creating a race of beings after their own image, with mechanisms rather like themselves or, instead, a race of beings made after the fashion of the more nervous subgroup of Martians. (They would not be able to control the roamings of such beings, so they would know in advance that some would be exposed to false informants, some to true informants.) Insofar as the dogmatic Martians had a choice about which kind of beings to create (without having subsequent control over them), we could well imagine that they would choose the mechanisms of the nervous Martians. Of course, this choice would depend on further judgments: if they judged that the value accruing to certain bits of knowledge generated by the dogmatic mechanism outweighed the chaos resulting from the inevitable negative performances of that mechanism, then they would still prefer a dogmatic mechanism to a nervous one.

Achieving epistemological stability is one thing. Thinking of one's internal mechanisms as a good prototype for an epistemological engine factory is quite another. While it may not induce internal epistemological turmoil, it is at least somewhat tragic were one to come to think of one's internal mechanisms as a disastrously unwise prototype for the factory. And even if one valued one's knowledge so much that one thought of one's internal mechanisms as an all things considered desirable prototype, it remains somewhat tragic that an all things considered desirable prototype would inevitably wreak so much irremediable havoc for so many of its manifestations.

Footnotes

¹ Very little in this essay is original. The material under ‘fine-tuning’ articulates points that I have picked up in conversation with Frank Arntzenius and (especially) David Manley, and much of the material under ‘Trust’ stems from conversations with Maria Lasonen and Timothy Williamson. (Not that any of them would approve of the particular spin that I have given to various points.)

² For further discussion see Paul Davies, *The Accidental Universe*, Cambridge University Press, 1982, Richard Swinburne, ‘Argument from the fine-tuning of the universe,’ in J. Leslie, ed., *Physical Cosmology and Philosophy*, Macmillan, New York, 1990, p. 154-173.

³ I have the standard Bayesian approach in mind

⁴ She will be similarly undisturbed if she thinks that a variety of initial credences are rationally permissible for that being, and that certain rationally permissible initial credences are weighted in the manipulator-friendly way.

⁵ See Alvin Goldman, ‘Discrimination and Perceptual Knowledge,’ *Journal of Philosophy*, November 1976. For further discussion, See Hawthorne and Gendler, ‘The Real Guide to Fake Barns: A Catalogue of Gifts for your Epistemic Enemies,’ *Philosophical Studies*, June 2005.

⁶ He wouldn’t put this point by saying that he is so constituted that the order of his evidence makes a difference to where he ends up. He will think that a case in which he read the false library first is a case where certain bits of evidence – for example *that P* – where P is something he learnt from the true library, would never have been acquired.