Wondering Why

Roger White

MIT

Draft for Rutgers Epistemology Conference

Need to add a bunch of references and stuff

We're curious about stuff. Curiosity drives inquiry. We start wondering whether P, or where X is, or what Fs are like, or why it is that Q, and then set about trying to figure things out. What is it to be curious? A first stab would be that it's wanting to know something. That seems on track but I'll start by raising a bunch of questions about it. Eventually I'll get to my main interest here which is a special kind of curiosity: explanatory curiosity. We wonder why about somethings and not others. I'm curious why we do this. Not just why we wonder why, but why we (should?) wonder why P but don't (needn't?) wonder why Q. I'll try to draw some connections between explanatory curiosity, rational inquiry, simplicity, symmetry, belief and credence, and other fun stuff.

Okay, so being curious is wanting to know something. Or so it seems. What do I want to know? Not just anything will do. Sometimes what will satisfy me is quite constrained: If I'm curious whether P, I need to either end up knowing that P or knowing that not-P. Nothing else will do. Other times there is a large class of propositions knowledge of any one of which I'd be happy with. But these might all fit a constrained form varying with some parameter. I'm wondering when Jennifer's talk starts. The knowledge I'm after will take the form *Jennifer's talk starts at t*. In other cases the class of satisfying items of knowledge is more open ended and heterogeneous. I'm curious why Trump won't release his tax returns. What I'm after is something of the form *Trump won't release his tax returns because P*. Here for *P* we plug in a proposition, or fact. But unlike times, facts are such a varied bunch. And I could hardly begin to specify the limits of some interesting class of serious candidates. The short take home here is just that curiosity is a state focused on a question. To be curious whether P, when X is, why Q, etc., is to want to know the answer to the question. The question determines a class of propositions, knowledge of one of which would satisfy the desire—although we might not always be able to say much about this class of propositions other than that they are potential answers to the question. I will assume that curiosity is directed toward a question. Although perhaps we can also be curious in a less directed way. I can be curious about physics without yet knowing which questions to ask.

But is it really *knowledge* that we are after? Here's an apparent reason to worry. Aren't there things we are curious about that we really don't want to know the answer to? You tell me you don't want to know how the last episode of *House of Cards* turns out and go out of your way to avoid spoilers. But you're still obsessively curious. Here the natural thing to say is that you have conflicting desires. You do want to know how it ends. After all you are going to binge watch the next five episodes tonight to find out. Your desire to find out by enjoying the show wins out over your desire to know now. A different sort of case involves the category now known as TMI. Perhaps you can't help but have some lurid curiosity about what so and so and what's his name are up to. But you realize it's none of your business and that if you stumbled on the facts you would recoil and wish you could unlearn it. Here the conflict might be that you want to know what's going on but don't approve of your desire to know.

We can distinguish here between *pure* or *intrinsic* curiosity and *instrumental*. You can be curious about something like the ending of a TV show that has no further practical or theoretical significance to anything else you care about. You're not going to *do* anything with the knowledge. On the other hand, you can wonder when the next conference session starts not because of any intrinsic interest but just because you want to get there, and without such knowledge you are unlikely to do so. Or you might be fascinated by mathematical conjecture C. You want to know whether lemma L is true, but only because it will allow you to prove C. L itself is of no intrinsic interest to you. The two can coincide of course. You are wondering what exactly are the relations between the Trump administration and Russia. The answer may have major practical upshot and figure into your political strategy. But it's also just bugging the hell out of you. But they can also come far apart. You can have strong instrumental reason *not* to know P while being very curious about it. A mind-reading psychopath announces "Anyone who knows whether 37 + 25 = 62 will be shot!" The wise course is not to think about the matter. But immediately many of us will be thinking "But does it equal 62?" Some of us will end up getting shot because our curiosity got the better of us. That bit of math wouldn't normally be that interesting to us. It's only because we strongly want not to know the answer for instrumental reasons that we find ourselves curious and wanting to know the answer.

But what's so great about knowledge that we should want it so much? Knowing involves meeting a bunch of conditions like perhaps having a true belief, being justified in holding it, not basing your belief on a false lemma, not being such that you could easily have falsely believed in on a sufficiently similar basis it or something sufficiently similar. Why not be satisfied with something weaker? Start with just a mere belief. I wonder whether Trump will be impeached. I realize that if I end up forming a belief, true or false on the matter I will consider myself satisfied and inquire no further into the question. But this doesn't show that the truth does not matter to me. Although the question may be bugging me, I have little interest in popping any belief-inducing pill to relieve my curiosity. Relief from the psychological tension of an unresolved question is nice but it isn't all I'm after. I want the truth. Of course if I do take a pill and thereby believe that Trump will be impeached I'll *think* that I've gotten what I was after since unless I'm confused I'll think that my belief is true. Indeed, I'll take myself to know it. (For the pill to do its trick it will have to erase my memory of having formed my belief this way, and give me the illusion of having based my belief on solid evidence. If I doubt that I've done so I'll naturally become skeptical). I won't find myself in a state of thinking, "Trump *will* be impeached. I have no idea if this opinion of mine is true, and I don't really care whether it is or not. But I'm glad to have settled the matter to my satisfaction."

So truth matters. But why ask for more than that? There are advantages to basing one's opinion on the best evidence available that go beyond merely getting the answer right in a single instance. Forming opinions in a good way gives me a better shot in general at getting things right. Poorly forming beliefs even if true is a bad habit to get into. And even in the individual case, items of knowledge tend to enjoy greater *resilience* than mere true beliefs, even ones that are justified. E.g., a justified true belief based on a false lemma can be dislodged by the discovery of the error.[1] Resilience of belief in practical matters can be advantageous. But while there are reasons to want knowledge over mere true belief or even justified belief, it is not so clear that these are involved in *curiosity* as such. I'm wondering what time it is. I look at the clock and correctly conclude that it's 10:35. Now I learn that the clock stopped a while

[1] Williamson 2000, Das (ms)

ago. I don't know that it's 10:35 after all. I'm disappointed as that puts me back into a state in which I have no idea what time it is and want to find out. But next I find that the clock stopped *exactly twelve hours ago*. I come to know that I've only just now come to know that it's 10:35 although my belief up until now was justified and true. I'll be amused to learn that I was just Gettiered. But should I conclude that my curiosity about the time was not satisfied after all, that I merely thought that it was (although now finally it is)? It's not clear to me that I should. I wondered what time it was. I concluded it was 10:35. I was right. It can seem like a mere amusing oddity that unbeknown to me at the time I arrived at the correct answer to my question in a quirky way. Suppose I could take an anti-Gettier pill that prevents one from getting into these situations. Given the points above about resilience there is some practical advantage to taking such a pill. But when it comes to pure curiosity it's not clear to me that it really matters to me whether my true belief counts as knowledge. A similar point could be made about justification. I learn that my belief that P is correct but was (until now) based on faulty reasoning. I may experience some epistemological shame for this. But as far as my curiosity concerning P goes, it's not clear that I've missed out on anything.

So perhaps all I'm after when I'm curious is *true belief*. Why then is it so natural to describe it as wanting to *know*? Perhaps the answer has to do with the way that true belief and knowledge are not separable from the first-person point of view of inquiry. Contrived cases aside, there isn't a way of seeking true belief which isn't equally a way of pursuing knowledge.[2] If I'm in the business of forming a true belief about whether P, the best I can do is to examine the evidence carefully and form my opinion rationally in response to it. But this is just as much a way of coming to know whether P. There is a lot more to knowing that mere true belief. But there isn't a way of aiming at less than knowledge but still at truth. Any method that lowers my chance of gaining knowledge likewise lowers my chance of being right. Similarly, when I do form an opinion I can't very well take myself to have arrived at the truth but not at knowledge. Any reason to doubt that my belief constitutes knowledge will be a reason to doubt its truth. So I can't reasonably come to the conclusion that I don't know whether P but that still I've got what I was after since my belief is true.

[2] Contrived cases include the taking of a *pro*-Gettier pill.

Here is another reason to question the link between curiosity and knowledge. I can be curious about matters where I have little hope of gaining knowledge. I may even know that I *can't* know whether P but still wonder whether it's true. I might wonder whether there is an unknowable God.[3] More precisely, I might wonder whether

UG: There is a God but no one can know that there is.

If knowing a conjunction entails being able to know it's conjuncts then the factivity of knowledge prevents me from knowing UG. Knowing the falsity of UG is not logically ruled out. I could know that UG was false if I knew there was no God. But it is hard to see how I could know this especially if the God there is might be an unknowable one. Such knowledge couldn't be grounded in my failure to find such a God since an unknowable God is hard to find. I could know that UG is false by knowing not only that there is a God but that I know there is, or at least that someone does. But my prospects for gaining such ambitious knowledge and knowledge of knowledge might seem slim. In any event, my interest is more in knowing that there *is* an unknowable God than that there isn't. If our predicament is such that there is a divine creator that is beyond epistemic reach then I should like to know this. But I know that *that* is knowledge I can't possibly have.

Of course it's possible to want something you can't possibly have. And it might even be rational to do so. I wish there were more hours in a day. Platonic solids are beautiful and elegant forms. I would like it if there were more than five of them. Perhaps knowing that there is a God whose existence will never be known would be a valuable state to be in, were it possible, as it would involve accessing a profound fact about our predicament. But can it be rational to *pursue* something you know you can't possibly have? And yet we do inquire into things where knowledge is not a serious option. I hope you still remember when most of us were obsessing over NPR's true crime podcast *Serial*. We stayed up thinking about whether Adnan committed the murder by noodling over the various items of evidence. What about the Nisha call? Could it have been the result of a butt-dial? And so on. But it was quite clear that the evidence was not sufficient to ground anything close to knowledge on the matter. We

[3] John Hawthorne suggested a case of this sort.

knew we weren't about to form an outright *belief* on the matter one way or another no matter how much we pondered the evidence. The most that was going to happen is that our opinion would *lean* a bit more in one direction or another than it did before. But we did nevertheless inquire into the question of who committed the murder. Was this just a futile exercise, like trying to square the circle, or invent a perpetual motion machine?

To make sense of what we are doing it seems we have to appeal to something besides beliefs. Credences are the obvious choice. I suppose someone might say that what we want to know is *how likely it is that Adnan did it*. But if this understood as an outright belief in a proposition about the probability of another proposition then it seems to get the subject matter wrong. Our interest is focused on whether Adnan did it. Questions about the probability given the evidence are secondary.[4] Credences don't have a truth value. But they can be more or less close to the truth, they can be more or less accurate. This framework of credences and accuracy measures has been very popular of late. Not everyone is a fan. But it seems to me that the current puzzle about curiosity and inquiry gives us an additional reason to appeal to this framework. Even if we have no hope of coming to a conclusion about whether P, we can aim to increase the accuracy of our credence on the matter. Even if what I would like most is to have a true belief on the matter, I will settle for having my credence in P increase of it's true and decrease if it's false. There is no guarantee of course that my credence that Adnan did it will increase in accuracy as I think through the evidence. When the evidence is that weak and sketchy it can very easily be slightly misleading and send me in the wrong direction. But still I should take it that forming a credence on an evaluation of my total evidence to maximize the expected accuracy of my credence.

While I think I've made some progress here in understanding curiosity, there is a respect in which what I've said so far does not seem to get to the bottom of things. On reflection, no condition of this sort—wanting to know, wanting to believe truly, wanting increased accuracy—is really sufficient for curiosity. Bill wants to know how many blades of grass are on his front lawn. In fact, he wants to know everything. He sees the possession of knowledge as a kind of ideal to be pursued. After all, God knows how many blades of grass are on every lawn.

[4] See Moss for an account that might fit well here.

And God is a kind of ideal being in various respects including cognitively. At the very least Bill wants his beliefs to be true. Like many philosophers he takes the *correctness* condition of belief to be truth. Following Anscombe he sees the direction of fit for beliefs is that beliefs are to conform to how the world is. But Bill really couldn't care less how many blades of grass are on his lawn any more than we do. He couldn't care less why the sky is blue, or how life arose, or whether the Axiom of Choice is true. He leads a cushy life and has no need even for practical knowledge. (All his needs are taken care of by a team of servants who work around the clock to satisfy his preferences). Bill is not remotely curious about anything. But his passion for accurate credence, for true belief, for knowledge, is unsurpassed. Bill is a kind of *epistemic fetishist*. He's a bit like the guy who visits a sick friend in hospital *because it's the right thing to do*, not because he gives a crap about him.

It feels like there's a kind of dilemma here. On the one hand it seems obvious that to be curious is to want something. After all, curiosity motivates inquiry. What else can move you to do something other than a desire? And the content of this desire must involve some kind of mind-world relation. The goal of inquiry is to get things right. But once we put things this way we seem to have missed the essence of curiosity altogether. Alice wonders whether there is life on other planets. Unlike Bill, her interest is directed at the *world*, not on some mind-world relation. Ask them, 'So, you are hoping for a certain relation between your mental states and facts about the universe?' Bill will readily reply *yes*. Alice might say, "There are indefinitely many relations my mind might stand it to the world. What the hell do I care which one is instantiated? What fascinates me is whether we are all alone here or if there is life on other planets." Alice might even be a Humean skeptic about the self. Or a Churchlandian eliminativist about mental states. These stances are doubtfully coherent. But none of this prevents her from being intensely curious. I'm not entirely sure how to resolve this. That's philosophy for you.

Explanatory Curiosity

Explanatory curiosity arises when we know that P but wonder why it is that P. We could ask this for any P we happen to know. Why is the third digit of my phone number even? Why were more people in this room born on a Wednesday than on a Thursday? But we don't. For the vast majority of propositions P, we don't care why it is that P or if there is any reason why. Here's a different case that I've liked for some time. When you blow some soapy water through a loop it forms a perfect sphere. Why is that? Interesting question. One that it's natural to be curious about well before having any inkling of the answer. We are not struck in the same way by just any blob of soapy water. I spill some on the floor here and it lands in some irregular pattern. There are lots of ways that it could have landed. We don't find ourselves asking why it landed in precisely *this* way. At least not with the same urgency as we do with the soap bubble. We are more content to say something like 'Well, it had to land somehow.' I've been interested in this distinction for some time and have argued that it plays a role in the way we reason in a number of different contexts such as enumerative induction. Now I want to take a few more steps in making sense of what's going on. Along my journey I keep coming up with counterexamples and counterexamples to my account, which can be annoying. Some counterexamples strike me as "deep", revealing that an account is on the wrong track or needs serious revision. Others are more fussy: The account seems of basically the right form even if it could do with some Chisholming. I need to get this paper written so I can't hope to get to the bottom of things. I'll have to content myself with being somewhat sketchy and programmatic. My hope is to draw some interesting connections between things and give a sense at how a thorough account will go.

Having given myself a little license for sloppiness, let's get to it. The phenomenon I'm talking about are those cases that are sometimes said to "cry out" for, or "demand" explanation. It's closely related to the idea that some facts are "surprising" or "puzzling" in a certain sense. One thing such cases have in common is they are unexpected. At a first pass then, a necessary condition of P's being puzzling in this sense is that prior to learning P I judged it to be very unlikely. This needs some finessing. Once I've seen soap bubbles form I'll expect it to happen again even if I don't know why. I might still like to understand why it has happened in this very instance even though it is what I expected because I've seen it before. Perhaps we should say that what is really puzzling is that soap bubbles form at all. Prior to learning that they do, that was highly unexpected. We expect the answer to satisfy us with respect to any particular instance. This low prior probability is to be understood as conditional on a certain subset of my beliefs. Which ones? For the most part I want to say that they those that concern matters potentially explanatorily relevant to the phenomenon. Here is the basic idea. I roll a die a hundred times and it lands on each side about equally often. Nothing surprising about that except that the sequence of rolls landed 1, 2, 3, 4, 5, 6, 1, 2, 3, 4, 5, 6, ... Prior to tossing I made certain assumptions about the die: that numerals 1 to 6 are painted on the six sides, that it is approximately symmetrically weighted, that each outcome has no causal impact on subsequent (or previous) rolls, … It might be hard to spell out all that belongs on this list. But I'm implicitly making some assumptions of this sort and it is on that basis that I assign a low probability to the sequence 1, 2, 3, 4, 5, 6, 1, 2, … It's somewhat important that it's not just any beliefs conditional on which that outcome is unlikely. A trusted time-traveler tells me that my next toss of this coin will land Heads. I believe her but she is wrong (she was watching a subsequent toss). I'm surprised that she is mistaken. And perhaps this will leave me wondering why she made a mistake. But this does nothing to make the fact that the coin landed Tails at all puzzling. Nothing remarkable about that. It could just as easily have landed Tails as Heads. The time traveler also told me that the die wouldn't land in a revolving ordered sequence. So my low expectation that it would do so was overdetermined. But her saying so has nothing to do with what makes that outcome puzzling or in need of explanation. While reports from time-travelers returning from the future are relevant to assigning probabilities to outcomes. But they play no *explanatory role*. However the die ends up landing, it won't have landed that way because the time traveler told me so. But it might be because of the weight distribution of the die, the way it was rolled, and so forth. It is natural to say at this point that the relevant asymmetry here is between cause and effect. The rolling causes it to land which causes the time traveler to report on it. This is about right and in what follows I'll be focusing on cases where the explanatory information is causal information. But I don't want to restrict it to this for the following reason. It seems that these ideas of surprisingness and need for explanation can arise in the realm of pure mathematics where nothing causes anything. I actually worry a lot about how much the ideas I'm pursue here can be carried over to pure math. But I'm going to ignore that issue for now.

Now of course improbability conditional on my causal assumptions is not enough to make it puzzling. As we all know, any sequence of a hundred die rolls is antecedently improbable. But most are unremarkable. I don't find myself puzzling over why die landed 6352464125363431214… I figure it had to land in some sequence and it could have been this one as much as any other. Take another familiar example. Case 1: A billion people buy a lottery ticket and Jane Bloggs wins. So what? We knew someone was going to win and Jane was one of those with a ticket. Case 2: There are three lotteries with a thousand tickets each. Jane wins all three. We are suspicious and wonder why she won all three. Case 3: You are betting against Jane on the outcome of coin flips. Jane wins thirty times in a row. This is astonishing unless there's some trickery going on. Her success demands some explanation.[5] The antecedent probabilities are the same in Cases 1 and 2, and about the same in 3. But the explanatory urgency varies dramatically.

Apart from antecedent improbability, what else does this explanatory urgency consist in? The natural thing to say here is that something calls for an explanation to the extent that we have reason to think it has one. In Case 2 we think there's likely to be a reason she won all three. In Case 3 there's *gotta* be a reason she keeps winning. In Case 1, well, she probably just got lucky. One trouble we face here is that this gets us into the problem of what counts as an explanation. Some philosophers think there's a reason for everything, that for any P there's an answer to the question *Why P*? But they don't find every improbable fact surprising. If determinism is true, then prior conditions together with the laws of physics entail who will win the lottery. Does that count as an explanation of why Jane won? If so, then no matter what happens there's an explanation for it. But we don't want to say that determinism entails that everything is puzzling and equally in need of explanation. We must have in mind a certain kind of explanation. When we say that there must be some reason that Jane keeps winning, we don't just have any kind of reason in mind.

Here's a stab at what's distinctive about the kind of explanation that matters here. It's an explanation that exhibits a certain kind of *stability*: It is not one that depends on a precarious set of conditions that could easily have failed to obtain. Why did Jane's ticket #847382748 win? We could trace the causes back to the way the precise way the tickets were arranged in the barrel, the specific set of perturbations that they underwent, and so on. This would be a massively complex story involving numerous causally independent variables. Given these precise conditions, Jane's ticket was guaranteed to win. But these conditions themselves could

[5] Cases 1 and 2 are from Horwich 1982 which puts the matter in terms of 'surprisingness' but not explanation.

easily have failed to obtain. If any one of numerous factors—the position of tickets *x, y, and z*, the momentum of the spinning barrel at time *t*, …—had been slightly different then a different ticket would have won. Of course we can trace the causal story back further to include the skeletal structure and muscular contractions of the fellow who picked up the tickets and dropped them in the barrel. Or we can go back further and down to the microphysical level and take the various positions and momenta of particles back at the big bang from which, let's suppose, we can derive with the deterministic laws how things will turn out in every detail. We are still appealing to a set of conditions that could so easily not have obtained. This kind of precarious explanation is not what we expect to obtain in those cases we think of as 'crying out' for explanation. Such an explanation does not satisfy us. Go back to our soap bubble. The actual explanation as you might expect is rather elegant. It has to do with the attractive and repellent forces between soap and water molecules forming a thin film. It bobbles around and reaches equilibrium in a perfect sphere as that is the shape with the smallest surface area for a given volume. There's more to the story, but the crucial point is that it exhibits a kind of explanatory stability. The explanans does not involve some precise set of initial conditions that had to be just right. Soap bubbles are a cinch to make. The molecules in the liquid and surrounding air can be in just about any arrangement. Blow on it a bit and bubbles will pop out. Without knowing much about how this could work, an explanation of this stable sort is just what we expect. Before blowing on some soapy water we might judge it exceedingly unlikely. We figure the only way something like that is going to happen is if some very specific set of initial conditions obtains. (Compare the case where we throw handfuls of sand in the air. It would take an extraordinary combination of coincidences for it to turn out that the sand forms into the surface of a sphere). But as soon as we see the bubble we rightly think that it must have come about in a way that was relatively easy. We think that there is some story to be told given which soap bubbles form in a way that is not at all sensitive to precise initial conditions. It turns out we are right. But we had every reason to expect this just by observing the phenomenon. When we find ourselves wondering why soap bubbles form we are looking to fill in a story of a form that we already expect to be there.

There is a connection here with the way we inductively extrapolate. One soap bubble is enough for us to expect that there will be more. We find it implausible that it was just some freak occurrence. Blow a few bubbles and we quickly become pretty confident that we will get more of the same even if we do not yet have much of a clue about why this is happening. Not so with the irregular blobs we drop on the floor. Any series of such blobs with have something is common. The first one has the specific blob shape 1, the next one blob shape 2, … They all have the property of either having blob shape 1, or blob shape 2, or…, or blob shape n. Call anything with this attribute a schblob. Noting that all our blobs so far have been schblobs, we don't project schblobhood onto the next one we drop. Why not? We think the only explanation for an item's schblobiness is an unstable one. It involves some very detailed set of prior conditions of the molecules in the water and the air and the way we tipped the cup such that without things just thus and so no schblob would have resulted. Of all the ways these conditions could be only a fraction are conducive to schblobdom so we have little reason to expect more of the same. Not so with our soap bubble. We think that whatever is going on with soap bubbles their occurrence must be compatible with a wide range of conditions and so we expect to see more of them without too much trouble.

In the cases that most urgently call for explanation it is only a story of this stable sort that we will find satisfying. An unstable explanation can lead to an unsatisfying explanatory regress. Consider again the sand we throw in the air. To our astonishment, at time *t* the grains form the surface a large perfect sphere before falling to the ground. WTF? It may well be that this was predictable from some prior conditions. Laplace's Demon kindly offers to help by noting that at time *t-1*sand grain 1 was at coordinates (x1, y1) with velocity vector v1, and g2 was at (x2, y2) and v2, and…and with the laws of physics this entails that at *t* they would form a sphere! I don't know that the demon has even succeeded in explaining why a sphere was formed at *t*. He's told us why g1 is at (x1', x2') at *t*, and why g2 is at (x2', y2') at *t*, and so on. And these facts together entail that they form a sphere. But I'm not sure that he has explained the latter fact. But in any event we will naturally respond, "Well fine, but *why* was g1 at (x1, y1) with velocity v1 at *t-1*, etc.?" We're not happy when the demon goes on, "Well at *t-2* g1 was at…" He could trace the causal story back to the big bang and specify some set of conditions S entailing with the laws that a sand particles thrown in the air will form a perfect sphere at *t*. It's still the case that S could very easily not have obtained. And the very fact that this set of conditions happens to be one of the very few that lawfully lead to the sand forming a sphere at *t* is part of what makes us want to still ask why condition S obtained. We don't get this kind of regress of unsatisfying explanations with the soap bubble. We feel like we've made satisfying sense of what is going on when we see how soap bubbles can form so easily.

So far I've suggested that something's calling for an explanation has to do with our having reason to suppose it has one of a special sort. Can we say more about when we have such a reason and why? Paul Horwich gives the following account of what he calls the 'surprisingness' of an event which I think maps on pretty well to what I'm after here.

> Let C be our belief about the circumstances that we initially took to obtain as E came about. E is surprising if $P(E|C)$ is very low, but there is some initially unlikely but not wildly improbable alternative hypothesis K concerning these circumstances such that $P(E|K)$ is high.

This fits pretty well with the account I want to give. The circumstances that Horwich appeals to I think will have to be what I was identifying as explanatorily relevant factors. But I don't think it can be quite right. Here's one of Horwich's cases. I toss what I take to be a fair coin a bunch of times and get all heads. The string of heads E is improbable conditional on C which includes the assumption that the coin is fair. There is an alternative hypothesis K—that the coin is heavily biased or double-headed—conditional on which E is to be expected. Of course any sequence of heads and tails is highly improbable conditional on C. Take an unremarkable sequence E': HTTTHHTHTTHHHTHTTHHHTTH. $P(E'|C)$ is very low. And we can easily cook up an alternative hypothesis K such that $P(E'|K)$ is very high. For example, perhaps an alien is controlling the coin and making it land in the order HTTTHHTHTTHHHTHTTHHHTTH. This generalizes, so it might look as though Horwich's account has the result that everything is surprising. Horwich blocks this result with the clause stating that K must be initially unlikely *but not wildly improbable*. The alien story is wildly improbable and so doesn't serve to render this boring sequence surprising. I'm not happy with this fix. It's not just that it's terribly vague what counts as wildly improbable. I don't see how we can adjust the threshold to make cases come out right. Perhaps it's unlikely that aliens are messing with me at all. But that doesn't seem to have any bearing on the surprisingness of the sequence. Even if I invited an alien over for dinner and encouraged him to play around with my coins, I still find the sequence

HTTTHHTHTTHHHTHTTHHHTTH unremarkable. Of course the hypothesis that an alien is controlling my coin is not enough to make it likely that we will get that sequence. We need a stronger hypothesis such as that he is controlling the coin *in such a way as to produce the sequence HTTTHHTHTTHHHTHTTHHHTTH*. And this is very unlikely as there are so many sequences of that length to choose from. On Horwich's account we will have to say that probability of an alien controlling the coin to get this very sequence counts as 'wildly improbable'.

My worry is with how this generalizes to other cases. When I threw handfuls of sand in the air we were very surprised to see them all form into the surface of a perfect sphere. I'm not sure what kind of alternative hypothesis K could render the sand-sphere highly probable. It is hard to think of any very substantive hypothesis that I could articulate or even entertain that could play this role. Perhaps I could hypothesize that aliens were messing with the sand. I don't know if that would count as wildly improbable. It is certainly *very* unlikely that aliens are up to anything around me. Conditional on aliens directing the sand grains perhaps it's not so terribly unlikely that they might move them to form a sphere. There are lots of other arrangements the aliens could choose. But arguably geometrically simple shapes like a sphere stand out as salient choices. Still, we can suppose that we have overwhelming scientific evidence that there are no aliens around. The evidence is so strong that the alien hypothesis counts as 'wildly improbable'. It remains very surprising that the sand grains formed a sphere. Laplace's demon could spell out a very detailed hypothesis concerning the precise positions and momenta of all the sand grains and air molecules conditional on which the sand-sphere is to be expected. But *this* hypothesis will have to be orders of magnitude lower in probability than the hypothesis above in which my alien friend made the coin land HTTTHHTHTTHHHTHTTHHHTTH as it contains vastly more independent contingent details. So it will also have to count as wildly improbable. It also seems that this hypothesis does not constitute an *alternative* hypothesis to my beliefs C about the circumstances in which the sand was thrown. I have various beliefs about how sand grains move when thrown, how they collide, and so forth, on the basis of which I assign a low probability to them forming a sphere when I throw them. But these assumptions seem to be entirely compatible with the demon's proposed hypothesis. His hypothesis is just one very specific and hence improbable way that these assumptions might be true. Lastly, we might try the logically weakest hypothesis that might play the role of K.

Perhaps something like: *there is something or other going on which makes it very likely that this pile of sand will form into the surface of a sphere when thrown in the air.* It still seems that we will have to say that this hypothesis is wildly improbable. Prior to throwing the sand I think it's super-duper unlikely to form a sphere. Far less likely than the alien's choosing to get the coin to land HTTTHHTHTTHHHTHTTHHHTTH. I'm going to have to think that it's extremely unlikely that there's anything going on that can make it likely. So it looks as though Horwich's account can't deliver the result that the coin toss sequence HTTTHHTHTTHHHTHTTHHHTTH is unsurprising while the sand sphere is very surprising.

My own account is very much in the spirit of Horwich's. The conditions on E's being surprising or calling out for explanation are just these

1. $P(E|C)$ is very low
2. $P(E|C) < P(E|\neg C)$

The simple upshot of 2. is that $P(C|E) < P(C)$. I start with certain assumptions about my environment, about sand grains, how they interact, how forces bear on them. I just have some kind of inchoate theory of the kinds of factors that potentially explain the behavior of the sand. Seeing the sand form sphere casts doubt on C. C rules all manner of alternative hypotheses. E is somewhat more to be expected on the disjunction of these. That's enough for E to be surprising and to raise the question of why it happened.

There is a heck of a lot more to be said about this. What do the cases in which this likelihood inequality holds have in common? Why should this lead us to have a special interest in explaining why it is that E? What is the connection between this and stable explanations? What does the *simplicity* of the the spherical shape have to do with our need to explain it? I can't get to the bottom of these things here. Perhaps another day.