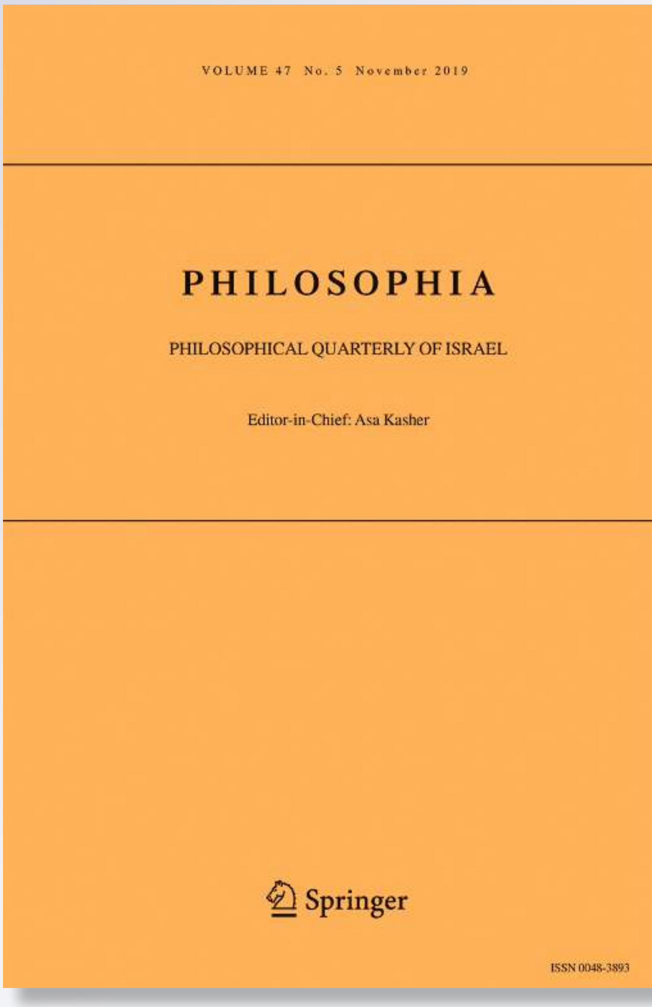
Why Do Certain States of Affairs Call Out for Explanation? A Critique of Two Horwichian Accounts

Dan Baras

Philosophia
Philosophical Quarterly of Israel

ISSN 0048-3893
Volume 47
Number 5

Philosophia (2019) 47:1405-1419
DOI 10.1007/s11406-018-0047-x



Springer

Philosophia (2019) 47:1405–1419
https://doi.org/10.1007/s11406-018-0047-x

Why Do Certain States of Affairs Call Out for Explanation? A Critique of Two Horwichian Accounts

Dan Baras [1]

Received: 14 August 2018 / Revised: 10 October 2018 / Accepted: 31 October 2018 /
Published online: 13 November 2018

Abstract
Motivated by examples, many philosophers believe that there is a significant distinction between states of affairs that are striking and therefore call for explanation and states of affairs that are not striking. This idea underlies several influential debates in metaphysics, philosophy of mathematics, normative theory, philosophy of modality, and philosophy of science but is not fully elaborated or explored. This paper aims to address this lack of clear explanation first by clarifying the epistemological issue at hand. Then it introduces an initially attractive account for strikingness that is inspired by the work of Paul Horwich (1982) and adopted by a number of philosophers. The paper identifies two logically distinct accounts that have both been attributed to Horwich and then argues that, when properly interpreted, they can withstand former criticisms. The final two sections present a new set of considerations against both Horwichian accounts that avoid the shortcomings of former critiques. It remains to be seen whether an adequate account of strikingness exists.

Keywords Calling for explanation · Strikingness · Surprise · Fine-tuning arguments · Arguments from design · Benaceraff-field arguments

1 A Tale of Two Dice

Many believe that it is implausible for certain states of affairs to be mere coincidences. According to this view, some states of affairs *call for explanation*. This idea serves as a premise in several influential debates in metaphysics, philosophy of mathematics, metaethics, modal theory, and philosophy of science. Despite its prevalence and importance, however, the distinction between things that call for explanation and things

✉ Dan Baras
dan.baras@mail.huji.ac.il

[1] The Martin Buber Society of Fellows, Mandel School for Advanced Studies in the Humanities, The Hebrew University of Jerusalem, Mt. Scopus, 9190501 Jerusalem, Israel

that do not has thus far received little careful attention. I will begin this paper by presenting the general idea that there are things that call for explanation. In the sections that follow, I will critically examine two dominant accounts of this distinction.

Imagine that you take a seemingly normal die, call this Die1, and, when you roll it 10 times, it lands in the following sequence: 1642346124. Given that, as far as you can tell, the die has six symmetrical sides, the subjective probability of the die landing in this precise sequence, prior to having tossed the die, is extremely low. Despite this exceedingly low probability, however, it probably does not strike you as calling for explanation. The die was bound to land one way or another, so why not this way? Now compare Die1 with a different die, call this one Die2. Like the first die, you role Die2 10 times, and it lands in the following sequence: 5555555555. As far as you can tell, Die2 looks like a symmetrical, six-sided die. Nevertheless, you would probably believe that the sequence is unlikely to be coincidental. Intuitively, the fact that the die landed on the same number so many times in a row should have an explanation. This is so even if it is difficult to explain why the reasoning from above does not apply here as well: That is, the die was bound to land in some sequence or other, so why not 5555555555? To be clear, sometimes we do consider chance a kind of explanation. Sometimes the chance explanation is even the correct explanation. However, in examples of this sort when people say that a state of affairs calls for explanation, they presumably mean to deny that the explanandum is the result of chance. Rather, they mean that it calls for a non-chancy explanation.

Based on examples such as these, it has seemed to some philosophers natural to conclude that there is an epistemically significant distinction between states of affairs that call out for explanation and states of affairs that we can more comfortably accept as accidental or brute. If so, the question arises. What is it that distinguishes the states that call for explanation from those that do not call for explanation? My working hypothesis for this paper is that there is some single property responsible for this difference. Following Schechter (2010, n. 33), I call the hypothesized property 'strikingness'.

The objective of this paper is twofold. One obvious objective is to argue against a popular account for strikingness. However, this is not my only objective. It is no less important for me in this paper to contribute to the clarification of the account and of the dialectic. Many of the claims I make in explicating the Horwichian view, such as that there are really two different views attributed to Horwich, that Horwich himself doesn't quite endorse either 'Horwichian' view, and that there are certain shortcomings of previous critiques, were previously unnoticed as far as I'm aware.

The plan is as follows: In the next section, I will review and make more precise the epistemic phenomenon at issue. Section 3 will present two influential accounts for strikingness inspired by the work of Paul Horwich. I am not the first to argue against the Horwichian accounts. However, section 4 argues that previous critiques fail because they overlook an ambiguity of the term 'striking'. Finally, in sections 5 and 6 I argue against each of these accounts in turn. I summarize my conclusions in the final section.

2 Clarifying the Question

Before I examine particular accounts, I would like to review and make more precise the epistemic phenomenon at issue. This explication is essential for the arguments that will follow. The idea described in the previous section can be formulated as a principle consisting of two parts:

(1) There is a distinction between states of affairs that are striking and states of affairs that are non-striking.
(2) To the extent that S is a state of affairs that is striking, if theory T (a conjunction of propositions) implies that S has no explanation, then that is an epistemic reason to either disbelieve T or disbelieve that S actually obtains.

I call the conjunction of these two claims the *Striking Principle* because states of affairs that call for explanation are often referred to as *striking* (White 2005, 2015; Schechter 2017, 2018). Of course, there's a whole array of adjectives used in the literature, such as 'surprising', 'suspicious', 'remarkable' or 'puzzling'. For present purposes, I assume these terms are more or less interchangeable in this context. I will stick for the most part to 'striking' for consistency, except for passages in which I quote Horwich who uses the term 'surprise' instead.

The Striking Principle (or something similar enough) is used by philosophers to argue for the existence of God, for the existence of multiple universes, for the inexistence of mathematical and moral facts and against our having knowledge of first order logic. Here are two sample quotes expressing something along the lines of the Striking Principle:

> Clearly there are some states of affairs that do cry out for a non-chancy explanation, for which it would be foolish to accept as a mere coincidence. (White 2007, 461)

> Ceteris paribus, it is a cost of a theory if it treats some striking phenomenon as merely accidental or otherwise inexplicable. (Schechter 2013, 220)

Similar claims can be found in Field (1989), Leslie (1989), van Inwagen (1993), Parfit (1998), Manson (1998), Price (2002), Schechter (2010, 2018), Enoch (2011), Nagel (2012), White (2015), Mulgan (2015) and others. How substantive the Striking Principle is depends on what precisely strikingness amounts to. If, for example, strikingness just means any reason whatsoever to believe that a particular state is non-coincidental, then the Striking Principle is trivial. However, the arguments that rely on the Striking Principle (two of which are described below) are relying on a more substantive interpretation of strikingness. So what precisely is strikingness? This will be our concern in this paper.

In order to prevent misunderstandings regarding the Striking Principle, I would like to emphasize a few more points. First, there can be states that we are currently unable to explain, but, given our background theory, we have no reason to believe that they have no explanation. Furthermore, there can in principle be states that have an explanation, but given epistemic or cognitive limitations of ours, we, human beings, will never be able to learn or articulate the explanation. These scenarios are not what the Striking Principle is about. Claim (2) talks about a theory that implies that a given state has no explanation, not about our knowledge of the explanation or ability to know or articulate it.

Second, the kind of reason implied here is a pro tanto reason. In some cases, we do learn that striking states of affairs are, in fact, accidental. It is likely that somewhere there will be a die that lands in an ordered sequence, such as a series of fives, accidentally. And it is likely that when this happens, we may have strong reasons to believe that the series was accidental. In such cases, whatever reason we have to believe that the given state of affairs has an explanation is outweighed by the reasons we have to believe that it is accidental.

Finally, the epistemic reason is supposed to be independent. By that I mean that it is irreducible to more familiar principles of epistemic support such as predictive success and simplicity. If indeed it will turn out that the inferences that rely on considerations of strikingness are really grounded in better understood and more fundamental considerations, then I believe it is unhelpful and misleading to use the calling for explanation lingo.[1]

As mentioned, a lot rests on the Striking Principle. I will give two prominent examples. The first example is a family of cosmological fine-tuning arguments: In order for a planet like Earth with living beings like humans to exist, the initial conditions and laws of the universe must be incredibly finely tuned. If there were any slight change in the rate of expansion of the universe after the Big Bang, in the strong nuclear force that binds atomic nuclei, or in the force of gravity that keeps our planet at just the right distance from the sun, our universe would turn into chaos. Several authors have argued that the fine-tuning of the universe for life is striking and calls for explanation. Some argue that the universe's need for fine-tuning supports theism (White 2015; van Inwagen 1993), others argue that it provides a reason to believe that our world is teleological in some non-theistic sense (Nagel 2012), and still others argue that it provides a reason to believe that our universe is one of many universes (Leslie 1989; Parfit 1998).

Another example appears in the philosophy of mathematics. Mathematical Platonists believe that there are mind-independent mathematical facts and also that many of us, especially mathematicians, have many true mathematical beliefs. Combining these two beliefs, mathematical Platonists are committed to a high correlation between the set of mathematical truths and our mathematical beliefs. Hartry Field (1989, 2005) argues that this correlation is striking and calls out for explanation. He further argues that mathematical Platonists have no explanation forthcoming and, thus, that mathematical Platonism should be rejected. More recently, philosophers have recognized that Field's argument applies to any theory that implies that we have knowledge of non-causal truths, such as non-naturalistic normative realism (Street 2006; Enoch 2011; Bedke 2014). Joshua Schechter (2010) goes as far as to argue that Field's reasoning challenges our knowledge of first-order logic.

Interestingly, at present, the Striking Principle is mostly posited as a premise in these arguments with little elaborate critical discussion. Several questions require research: First, is the Striking Principle reasonable? Second, what is it that distinguishes states of affairs that call for explanation from those that do not? And third, when a state of affairs calls for explanation, what kind of explanation does it call for?[2] It is a working hypothesis of this paper that the striking principle is reasonable and, as already mentioned, that there is a

[1] For an elaboration of this point, see Baras, D., A strike against a striking principle, (unpublished).

[2] Possible responses to the third question are examined in White (2005) and Baras (2017).

single property, *strikingness*, that distinguishes states that call for explanation from those that do not. This paper will focus on proposed answers to the second of these questions and will carefully examine an influential account of strikingness adapted from Paul Horwich's work. In the course of my arguments, it will be important to remember the role that the striking principle plays in the arguments described in this section.

3 The Horwichian Accounts

In looking for an account of strikingness, a number of authors have appealed to a suggestion put forward by Paul Horwich (1982, 100–104) in a different context. Horwich was concerned with the idea that predicting surprising phenomena yields greater confirmation of a theory than predicting unsurprising phenomena. Horwich ends up arguing against this idea and thus his own view seems to be that surprises have no epistemic significance. Along the way, Horwich offers an account of what it is for phenomena to be *surprising*. Several authors have applied Horwich's account to the striking principle, suggesting it as a candidate account of *strikingness* in the context of the striking principle. Although I will begin with quotes from Horwich, my main discussion is directed at the later authors inspired by Horwich and not at Horwich himself.

Horwich begins with the observation that surprising explananda are highly unlikely given our background beliefs. For instance, according to Horwich, part of what makes Die2 surprising is that it is very unlikely that a fair die will land on 5 ten times in a row. However, as the opening examples demonstrated, low probability is not a sufficient condition for surprise. Horwich therefore proposes:

> The truth of E is surprising only if the supposed circumstances C, which make E seem improbable, are themselves substantially diminished in probability by the truth of E… that is, if there is some initially implausible…alternative view K about the circumstances, relative to which E would be highly probable. (Horwich 1982, 101–2)

There really are two distinct suggestions here. The distinctness of these suggestions has previously gone unnoticed.[3] Returning to the language of 'strikingness', the first suggestion is that something is striking if it calls into question background beliefs regarding the circumstances. The second is that something is striking if there is some competing hypothesis that, if it were true, would explain the given explanandum. (Shortly, I will defend my claim that these are indeed distinct views).[4] Hereafter, I will call the first view *disconfirmation*, since the idea is that some background assumptions

[3] For instance, Good (1984, 164), Schlesinger (1991, 99), Manson (2003) and Mogensen, A. L., Ethics , evolution , and the coincidence problem : A skeptical appraisal, (unpublished) attribute *competing hypothesis* (the first account that I will introduce shortly) to Horwich and make no mention of *disconfirmation* (the second account). Manson (1998) and Bostrom (2002, 30) both slightly misquote Horwich, and each of them attributes to Horwich a slightly different mixed account. Bostrom's interpretation of Horwich is logically equivalent to *competing hypothesis*.

[4] In private correspondence, Horwich explained that *disconfirmation* is his official account, whereas *competing hypothesis* was meant to be just one unnecessary way in which the account can be realized.

are disconfirmed, and I'll call the second view *competing hypothesis.* Both ideas can be illustrated using the opening examples. Why is 5555555555 striking? Perhaps because the sequence of fives puts into doubt our original belief that the die is fair (disconfirmation), or because it is nicely explained by a competing hypothesis that the die is weighted toward landing on five (alternative hypothesis). The sequence 1642346124, on the other hand, is not striking because no such alternative hypothesis is suggested, and we have no reason to doubt the background assumption that the die is fair. *Disconfirmation* is adopted by Bartholomew (1984, 46–47) and White (2000, 270, 2005, 3) as an account of strikingness. *Competing hypothesis* is adopted by Leslie (1989, 10),van Inwagen (1993, 135)[5] and Manson (1998, 2003).

I will make a few clarifications regarding these accounts. First, the accounts, or at least some natural interpretations of them, are not logically equivalent. A state of affairs E can disconfirm background beliefs without there being any competing hypothesis relative to which E is highly probable. It is true that in such a case there must be a competing hypothesis relative to which E has a higher probability than it does relative to C. However, the probability need not be high.[6] Therefore, a state can satisfy the conditions of disconfirmation without satisfying the conditions of competing hypothesis. Conversely, if we allow improbable hypotheses to count as competing hypotheses (see discussion in section 6), then it is possible for a state E to be potentially explained by a competing hypothesis without it being the case that E significantly disconfirms background beliefs.[7] In addition, I believe we should be open to the possibility that C can be disconfirmed even if we have no alternative yet in mind. The Bayesian framework does not allow us to attach posterior probabilities to hypotheses for which we did not already have some prior probability. In effect, the Bayesian framework is an inadequate model for situations in which we have not yet come up with a plausible competing hypothesis but we have some justified confidence that eventually we will. It therefore might be possible for some E to significantly disconfirm C without there being any known competing hypothesis that predicts E, merely because we are justified in expecting there to be such a hypothesis.

Second, following Horwich, I will sometimes formulate the account probabilistically. However, we should keep in mind that the Bayesian framework makes some idealizations that are not always appropriate in this context. I already mentioned the assumption that we have prior probabilities for all possible hypotheses. In addition, the Bayesian framework models rational credences, which are the end result of many reasons for belief figuring in rational deliberation. The Bayesian framework is not as neatly fit to analyze individual pro tanto reasons for belief. As noted before, the striking principle as I construe it is a principle regarding pro tanto reasons. Some of my

[5] In later editions van Inwagen revises his principle without explaining why. The existence of a potential explanation remains a core feature in his most recent revision (van Inwagen 2015, 205). However, the revised principle is presented as only a sufficient condition and it introduces new elements, some of which require disambiguation. Unfortunately, I cannot discuss van Inwagen's revision as it would lead us too far astray from the focus of this article.

[6] Example: Suppose there is only one competing hypothesis K, so that $K \equiv \neg C$. Suppose that (E) = 0.001, $P(C|E) = 0.5$ and $P(C) = 0.9$. Plausibly, these numbers fulfill Horwich's conditions for strikingness. They imply though that $P(E|\neg C) = 0.005$. Surely that doesn't count as high.

[7] Example: Again, for simplicity, suppose there is only one competing hypothesis K, so that $K \equiv \neg C$. Suppose further that $P(C) = 0.999$; $P(E|K) = 0.9$; $P(K) = 0.001$ and $P(E) = 0.01$. It follows that $P(C|E) = 0.91$. Arguably, that does not count as a significant decrease in the probability of C.

arguments will address the probabilistic formulation but I do not mean to rely too heavily on the Bayesian framework. I will therefore complement the probabilistic arguments with arguments that address the underlying informal ideas of these accounts as well.

Third, as the two accounts stand, they present strikingness as an all-or-nothing property. I do this for the sake of simplicity. It is more plausible to think of strikingness as a graded notion, i.e. that things can be more or less striking and call out for explanation to a greater or lesser degree. It is not difficult to revise the two accounts to accommodate a gradable notion of strikingness. For instance, regarding disconfirmation, we may want to say that the lower P(E), or the larger the gap between P(C) and P(C|E), the more striking E is. Something similar can be said for competing hypothesis. The arguments that follow are not affected by this simplification.

In the next section, I go into a short detour, explaining why previous critiques of Horwich's accounts are unsuccessful and drawing a general lesson about the relationship between the everyday notions of surprise and striking, and the technical notion of strikingness that we are concerned with here.

4 Why Previous Critiques Fail

There is a distinction to be made between the epistemically significant property 'strikingness', which we're seeking an account for here, and the everyday psychological notions of 'striking' and 'surprise'. A state of affairs can be surprising or striking in the psychological sense, without giving us reason to believe that the state is non-coincidental. If I were to win a million dollars in a lottery, I'd be very surprised (and find it striking). Still, I'd have no reason to believe that it wasn't a coincidence that I won. Conversely, a state can be unsurprising (or non-striking) psychologically, but nevertheless call for explanation. For example, if a coin has already landed a thousand times straight on heads, I will not be surprised if it continues landing heads. Still, I would have no less reason to believe that the fact that it continues to land heads must have an explanation.

As I'll now explain, previous critiques of Horwich's accounts seem to have overlooked this distinction. Schlesinger (1987, 222, 1991, 101) has criticized Horwich's account, and Harker (2012) has recently made a similar critique.[8] They reason as follows: Horwich's account implies that when all alternative hypotheses are conclusively ruled out and the initial background beliefs remain highly probable, the sense of surprise or strikingness disappears. Yet, it seems that ruling out alternative hypotheses does not vanquish our surprise. Applying their point to our opening examples, imagine that prior to tossing Die2, we carefully examine the die and accumulate conclusive evidence that the die is fair.[9] If you like, imagine an epistemic oracle telling you in advance that the tosses of Die2 are perfectly fair and the tosses independent. Suppose that we then rolled Die2 and it landed on five 10 times in a row. Would we be any less surprised? If anything, we would probably be even *more*

[8] For a discussion of Schlesinger and Harker's alternative accounts of strikingness, see Baras, D., Do extraordinary types call out for explanation?, (unpublished).

[9] Schlesinger and Harker use different examples to make the same point.

surprised. Schlesinger and Harker rightly argue that something can remain surprising even if we remain confident in our beliefs regarding the circumstances and rule out any alternative hypothesis.

However, to the extent that we are concerned with strikingness, in the technical sense, Schlesinger and Harker's critique misses the mark. In the case just described indeed we would be surprised in the psychological sense. However, it's not the case that we should believe that the sequence of fives has an explanation. Given that we are confident that the die and tosses are fair and independent (that's an essential feature of the case), we should conclude that the sequence of fives was coincidental, just as the Horwichian accounts of strikingness predict. Once we clearly distinguish the psychological notion of strikingness from the notion that appears in the striking principle, Schlesinger and Harker's arguments become less plausible.[10] The lesson is also important for the discussion that follows. It will be important to keep in mind the relevant notion of strikingness, namely, the property in virtue of which states of affairs call for explanation.

Because Harker's objection to Horwich can be resisted, and Harker's own account faces difficulties, Horwich's accounts remain the more attractive. My task in the next sections is to argue that there are nevertheless other reasons to reject Horwich's accounts.

5 Against Disconfirmation

Following Horwich, the following is a probabilistic formulation of the first account:

Disconfirmation: E is striking for an agent with initial credences P iff and because

1. The prior probability of E is low: $P(E) \approx 0$

   And,
2. Upon learning that E, a conjunction of beliefs regarding the circumstances C is significantly disconfirmed: $P(C|E) \ll P(C)$

As explained in the previous section, previous critiques of this account have attempted to provide counterexamples that demonstrate a mismatch with our intuitions about which things call for explanation. My objection to disconfirmation will be of a different kind. I will argue that the reasoning that led to the disconfirmation account is inconsistent with the reasoning that led to the acceptance of the striking principle to begin with.

The distinction between states of affairs that call for explanation and states of affairs that do not call for explanation is motivated by examples of the kind mentioned in my introduction above. Let us examine the example more closely. True, the sequence 5555555555 gives rise both to a judgment that the sequence must have an explanation, and to a judgment that relevant assumptions regarding the circumstances are called into question. However, we must ask exactly why these background assumptions are called into question. A natural answer is that they are called into question because they imply that the sequence is coincidental, whereas we have reason to believe that in fact it is

[10] Manson (1998) makes a similar point.

non-coincidental. The background assumption that is called into question is that the die is fair, and it is called into question because sequences of tosses of fair dice are necessarily coincidental, that is just what it means for a die to be fair. If this is correct, the background assumptions are called into question by the sequence of fives, because the sequence of fives calls for explanation. We should not therefore infer that the sequence of fives calls for explanation because background assumptions are called into question. The disconfirmation account loses its original motivation and is therefore unmotivated and loses its plausibility.

This problem with disconfirmation can be described as a Euthyphronic dilemma. Either a state of affairs calls for explanation because it calls our background beliefs into question, or our background beliefs are called into question because the state of affairs calls for explanation, but not both. The disconfirmation account sides with the first horn, while the whole line of thought that motivated the Striking Principle, and the arguments in metaphysics premised on it, side with the second horn.

Furthermore, disconfirmation implies that the arguments premised on the Striking Principle are either circular or ungrounded. Take Field's reliability argument for example. The argument relies on the premise that the correlation posited by Mathematical Platonism is striking. According to Disconfirmation, in order for the correlation to be striking, it must call background beliefs into question. On the basis of what could it call background beliefs into question? The basis cannot be that it calls for explanation and is unexplainable, because that would be circular logic. In order to avoid this circularity, there will have to be some other way that the correlation calls background beliefs into doubt. However, Field has not provided such a reason. Therefore, disconfirmation implies that the argument rests on an ungrounded premise. The same is true for van Inwagen and White's arguments that the fine-tuning of our universe cannot be coincidental (and therefore theism is supported). The same will be true for any other argument that wishes to question background beliefs on the basis of the Striking Principle. If disconfirmation is the correct account of strikingness, then no such argument is justified without a prior reason to doubt a theory. These authors do not provide such a reason.

I conclude therefore that if the Striking Principle is correct, disconfirmation cannot be the correct account of strikingness.

6 Against Competing Explanatory Hypothesis

Let us now consider Horwich's second account:

Competing hypothesis: E is striking for an agent with initial credences P iff and because

1. Prior to learning that E, E had very low probability: $P(E) \approx 0$

   And,
2. There is some hypothesis K among the set S of competing hypotheses regarding the circumstances, upon which E is probable: $\exists K \in S$: $P(E|K)$ is high

The account seems intuitive when we think of dice tossing sequences such as 5555555555. In such a case there is a salient and somewhat probable explanation for the sequence, namely that the die is weighted. However, what if a die was tossed 10 times and landed 3535353535? Intuitively, this sequence calls for explanation no less. But what could the competing hypothesis possibly be in this case? You might suggest: One competing hypothesis is that a magician is able to control the die, and this particular magician likes this very orderly sequence. However, if this counts as a relevant hypothesis, then it seems we can come up with a similar hypothesis for any sequence of dice tosses. The sequence 1642346124, which I took as a paradigm for a sequence that does not call for explanation, would be explained by the competing hypothesis that a magician with a special liking of this particular sequence is controlling the die.

The lesson is that there must be restrictions on the set (S) of propositions that count as relevant alternative hypotheses. Otherwise, the account will wrongly imply that any improbable state of affairs calls for explanation because the second condition is always fulfilled. Is there any plausible, non ad hoc restriction that can give the correct results? I will now examine two suggestions for such a restriction: 1. Hypotheses with non-negligible prior probability. 2. Salient hypotheses. I will argue that both variants of the competing hypothesis account face problems.

Perhaps the most natural suggestion is the one made by Horwich himself, who restricts the set of alternative hypotheses to those with initial non-negligible probability.[11] In this account, the first element to note is that it is far from clear how the terms 'non-negligible probability', '≈0', and 'is high' are to be interpreted numerically. The problem is particularly troubling because what would seem like very intuitive ways of interpreting these terms would imply that it is impossible for any state of affairs to fulfill the two conditions. Let me explain: Suppose we define these expressions as follows: Near zero means lower than 0.001; non-negligible means higher than 0.01; and high probability means higher than 0.2. At least prima facie, it seems plausible that if these numbers are incorrect, that is because near zero should be lower and non-negligible and high-probability should be higher. However, given these numeric definitions, it is probabilistically impossible for any state of affairs to be striking.

Here's why: The standard analysis of conditional probability and the axioms of probability imply that $P(E|K)\leq\frac{P(E)}{P(K)}$.[12] Suppose that P(E) = 0.001; P(K) = 0.01. It follows that $P(E|K)\leq 0.1$, i.e. that P(E|K) cannot be high. Things are only made worse if we revise our numeric definitions in the directions suggested above. In order for there to be anything that this account predicts to be striking, the numerical interpretations must be different from what the terms initially suggested. In particular, the upper bound

[11] Good (1956) also has this restriction. An interesting consequence of the restriction to non-negligible hypotheses is that necessarily anything satisfying this version of Alternative Hypothesis will also satisfy Disconfirmation, because E will raise the probability of K, which is contrary to a relevant background belief. The converse is false because E can disconfirm C even if all alternative hypotheses have negligible prior probability, for instance, if P(C) ≈ 1.

[12] Here's the proof:

1. $P(E|K) = P(K|E)\frac{P(E)}{P(K)}$ [Bayes' theorem]
2. $P(K|E)\leq 1$ [probability axioms]
3. $P(K|E)\frac{P(E)}{P(K)}\leq\frac{P(E)}{P(K)}$ [from 2]
4. $P(E|K)\leq\frac{P(E)}{P(K)}$ *QED* [from 1, 3]

of near zero must be closer to, and perhaps even higher than the lower bound of non-negligible.

Note though that I have been using Horwich's formulation according to which the first condition is that P(E) ≈ 0. The problem is mitigated if we replace the first condition with Harker's version, P(E|C) ≈ 0, where C stands for a relevant set of beliefs regarding the circumstances. Presumably, if P(E|C) is extremely low, then so will P(E), assuming that part of what it means for C to be a set of beliefs is that it is attributed high probability. Nevertheless, the argument from above will not run smoothly anymore. When the two conditions are satisfied, P(E) can be slightly higher than P(E|C), close enough to P(K) to remove the problem from above.[13]

Even if we avoid this first problem by revising the first condition or reinterpreting the terms, further issues arise. Recall that the striking principle was supposed to provide reasons for belief that are independent of other familiar principles such as simplicity and predictive success. However, if we accept competing hypothesis and limit alternative hypotheses to those that have initial non-negligible probability, then the background beliefs are disconfirmed without bringing in any of the sort of reasoning involved in the kinds of examples used to motivate the striking principle. Here is how. Suppose that there is some K such that P(K) is low but non-negligible and P(E|K) is high. Bayes' theorem, which is not based on considerations of strikingness, tells us that: $\frac{P(K|E)}{P(K)} = \frac{P(K|E)}{P(K)}$. Given that P(E) is very low and P(E|K) is high, it must also be the case that P(K|E) is significantly higher than P(K) for the equation to be true. In other words, E strongly confirms K. To illustrate, suppose P(E|K) = 0.9, P(K) = 0.01 and P(E) = 0.01. It follows that $P(K|E) = \frac{P(E|K)P(K)}{P(E)} = \frac{0.9 * 0.01}{0.01} = 0.9$. E significantly raises the probability of K, from 0.01 to 0.9. Since K is confirmed, and K is a potential explanation of E, we thereby have strong reason to believe that E has an explanation. This version of the competing hypothesis account will give the correct predictions,[14] but considerations of strikingness fall out of the picture.

Moreover, this version of competing hypothesis implies that we never just have reason to believe that some potential explanandum has an explanation. Rather, we can have reason to believe that a potential explanandum has an explanation in virtue of having reason to believe that a particular hypothesis K is correct.[15] This too is contrary to the Striking Principle that states that we can have an independent reason to believe that a state has an explanation. Take Die2, for example. Trivially, if we have reason to believe the particular explanation that Die2 is weighted towards landing five we thereby have reason to believe that the sequence of fives has an explanation. However, this is not an independent reason to believe that Die2 has an explanation, but a deduction from the fact that we have reason to believe that the die is weighted.

If an account of strikingness along the lines of competing hypothesis is to escape these charges, then it must not restrict the set of competing hypotheses to hypotheses

[13] I thank Blake McAllister for pressing me here.

[14] Notice though that in order for this account to give the correct predictions, the terms 'low probability', 'non-negligible probability' and 'high probability' should be interpreted not as fixed ranges of numbers, but rather as relative to each other. For example, the lower the prior of E, the lower K can be and still be confirmed by E. I thank an anonymous referee for this comment.

[15] This seems to be what Laplace (1902, 16) and Urbach (1992) believe; therefore, I do not consider them proponents of the Horwichian account of strikingness.

with non-negligible probability.[16] I now move to examine a second suggested restriction on the set of competing hypotheses. In a brief note, Joshua Schechter (2010, n. 33) suggests that the relevant alternative hypotheses should be restricted to salient hypotheses.[17] Plausibly, even hypotheses with negligible initial probability can be salient. For instance, if you were to hide a fallen tooth beneath your pillow at night and find in the morning a dime instead of the tooth, a salient hypothesis may be that a tooth fairy has paid you a visit. The tooth fairy hypothesis could be salient even if you initially attribute exceedingly low probability to the tooth fairy hypothesis.[18] However, this account runs into a different set of problems.

First, unfortunately, Schechter does not develop his idea, and no explanation is given as to what 'salience' means in this context. Presumably, the relevant sense of 'salience' is not the psychological one, meaning that something is salient if and because it is prominent in one's awareness. The psychological notion renders salience (and everything that depends on it) entirely contingent on epistemically irrelevant aspects of one's psychology. Neither can 'salience' be taken to mean 'the set of hypotheses that one ought to take into consideration,' because then we will have made no progress in discerning precisely which hypotheses one ought to take into consideration. So long as we lack an account of salience, it is difficult to judge the plausibility and explanatory value of this account. And so long as we lack an account of salience, it is difficult to understand what this account even says, and it seems like little progress has been made.

Another problem with the account is that it implies that a totally implausible hypothesis can give you reason to believe that something has an explanation, even though you should not at all believe the hypothesis. Consider for example the following possible scenario[19]:

[16] For this reason, Harker's suggestion to Horwich is ruled out as well. Harker (2012, n. 1) suggests that we 'limit admissible alternative circumstances to those that we might reasonably expect to know or discover.' However, we can only reasonably expect to know or discover hypotheses that we can reasonably expect to be true. Harker's suggestion fares even worse than Horwich's because we can easily come up with counterexamples. Suppose that an omniscient god revealed to you that you have no way of knowing or discovering the explanation for why a particular die landed 3535353535. Alternatively, suppose that the die fell into the ocean, so you cannot reasonably expect to be in a position to examine it ever again. It would be unreasonable to expect to know or discover any alternative hypotheses in such circumstances, yet the sequence seems no less striking.

[17] Although their wording is vague, it seems that Leslie (1989, 10) and van Inwagen (1993, 135) think along similar lines. Leslie indicates that the potential explanation must be a 'tidy' explanation. However, what does tidiness mean in this context? The same problems that I raise for Schechter's restriction to salient hypotheses apply to Leslie's restriction to tidy hypotheses so I do not see a need to discuss Leslie's idea separately.

[18] Adherents of this account may want to allow the set of alternative hypotheses to include hypotheses with initial probability zero, since, plausibly, hypotheses that are conclusively ruled out can still be salient. Even if the tooth fairy hypothesis were conclusively ruled out, one might still believe that the replacement of the tooth calls for explanation *because* of the tooth fairy hypothesis. If so, we will have to make some amendments to the formal account because conditionalizing on probability zero propositions is undefined in the standard Bayesian framework. Note that some philosophers believe that the lack of definition of conditional probability for probability zero propositions is a flaw in the standard Bayesian analysis of conditional probability (Hájek 2003).

[19] White (2007) makes this point. He argues that if we rule out the intelligent design hypothesis, we no longer have reason to rule out the hypothesis that we owe our existence to mere chance. This argument seems to me to be in tension with White's claims elsewhere in which he seems to imply that something can call for explanation even if the most salient explanations have been ruled out (White 2005, 2015).

I. A die lands 613251436253
II. A salient hypothesis for S is that K: = there is a demon particularly fond of the sequence 613251436253 that took control of the die.
III. $P(K) = 10^{-100}$

The salient hypothesis version of competing hypothesis implies that in these circumstances, S has reason to believe that the sequence 613251436253 has an explanation. This seems like a very implausible result. How could such an incredibly improbable hypothesis have such a significant epistemic implication? How could salience matter in such a case?

In sum, without a restriction on which competing hypotheses are to be considered in order to fulfill the second condition of competing hypothesis, the account collapses into the first condition. If the set of competing hypotheses is restricted to hypotheses with non-negligible initial probability, then in order for it to be possible for anything to be striking, the terms used to formulate the account must be interpreted differently than their intuitive meanings. Even if this issue is set straight, the account might give the correct result, but for the wrong reason. On the other hand, the suggestion that we restrict the set of alternative hypotheses to the set of salient hypotheses is too vague to be useful and has an unintuitive implication. As long as we lack a plausible candidate for how the set of competing hypotheses ought to be restricted, the success of the second Horwichian account of strikingness remains highly questionable.

7 Conclusion

I began this paper by defining a very influential epistemic principle that I have named the Striking Principle. According to the Striking Principle, there are states of affairs that we have strong reason to believe have an explanation. I then presented two logically distinct accounts of strikingness. Both are inspired by Paul Horwich's work and have impressed a number of authors. I argued that Harker's assault on these accounts is disarmed once we distinguish between the psychological concepts of surprise and striking and the technical concept of strikingness that interests us here. Finally, I raised a series of new problems with each of the accounts.

If I am correct and Horwich's accounts fail as accounts of strikingness, how shall we progress from here? One possibility is to search elsewhere for a better account of strikingness. If such attempts fail, we might have to conclude that no such account exists and either strikingness is unanalyzable or perhaps the Striking Principle is not a reasonable principle to endorse. I intend to explore these possibilities in future work.

What are the ramifications of this paper for the fine-tuning and reliability arguments mentioned in the introduction? Not much, I submit. Even if the Horwichian accounts fail, that does not yet imply that the Striking Principle is false, nor does it imply that the fine-tuning of our universe or our reliability in mathematics are not striking. What we can say at this stage, though, is that as long as we lack an adequate understanding of strikingness, it is difficult to assess the plausibility of the premises of these arguments.

Acknowledgements Whatever merits this paper may have, they are most definitely explainable: Very helpful comments I received from Ron Aboodi, Sharon Berry, David Enoch, Yehuda Gellman, Alan Hájek, Paul Horwich, Ofer Malcai, Neil Manson, Blake McAllister, Eli Pitcovski, Joshua Schechter, Miriam Schoenfield, Orly Shenker and Martin Smith on previous drafts have contributed significantly to improving this paper. I also wish to thank the participants at my presentations at Ben Gurion University's philosophy department colloquium, the 2016 Eastern Regional Meeting of the Society of Christian Philosophers at Rutgers University and at the 2nd Jerusalem-MCMP Workshop on Explanatory Reasoning in the Sciences at the Munich Center for Mathematical Philosophy for very helpful discussions. During the various stages of writing, my work was supported by Ben Gurion University and later by the Center for Moral and Political Philosophy at the Hebrew University of Jerusalem.

References

Baras, D. (2017). Our reliability is in principle explainable. *Episteme, 14*(02), 197–211. https://doi.org/10.1017/epi.2016.5.

Bartholomew, D. J. (1984). *God of chance*. London: SCM.

Bedke, M. S. (2014). "No coincidence?" In *Oxford Studies in Metaethics, 9*, 102–25. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198709299.003.0005.

Bostrom, N. (2002). *Anthropic bias: Observation selection effects in science and philosophy*. New York: Routledge.

Enoch, D. (2011). *Taking morality seriously*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199579969.001.0001.

Field, H. (1989). *Realism, mathematics, and modality*. Oxford: Basil Blackwell.

Field, H. (2005). Recent debates about the a priori. *Oxford Studies in Epistemology, 3*, 69–88.

Good, I. J. (1956). The surprise index for the multivariate Normal distribution. *The Annals of Mathematical Statistics., 27*, 1130–1135. https://doi.org/10.1214/aoms/1177728079.

Good, I. J. (1984). A Bayesian approach in the philosophy of inference. *British Journal for the Philosophy of Science., 35*, 161–166. https://doi.org/10.1093/bjps/35.2.161.

Hájek, A. (2003). What conditional probability could not be. *Synthese, 137*(3), 273–323. https://doi.org/10.1023/B:SYNT.0000004904.91112.16.

Harker, D. (2012). A surprise for Horwich (and some advocates of the fine-tuning argument (which does not include Horwich (as far as I know))). *Philosophical Studies, 161*(2), 247–261. https://doi.org/10.1007/s11098-011-9732-3.

Horwich, P. (1982). *Probability and evidence*. Cambridge: Cambridge University Press.

Laplace, P.-S. (1902). *A philosophical essay on probabilities*. Translated by Frederick Wilson Truscott and Frederick Licoln Emory. 1st ed. New York: John Wiley & Sons.

Leslie, J. (1989). *Universes*. London: Routledge.

Manson, N. A. (1998). *Why cosmic fine-tuning needs to be explained*. Syracuse: Syracuse University.

Manson, N. A. (2003). Cosmic fine-tuning, 'many universe' theories and the goodness of life. In W. B. Drees (Ed.), *Is nature ever evil? Religion, Scinece and value*. Abingdon: Routledge.

Mulgan, T. (2015). *Purpose in the universe*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199646142.001.0001.

Nagel, T. (2012). *Mind and cosmos*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199919758.001.0001.

Parfit, D. (1998). Why anything? Why this? *London Review of Books, 20*(2).

Price, H. (2002). Boltzmann's time bomb. *The British Journal for the Philosophy of Science, 53*(1), 83–119. https://doi.org/10.1093/bjps/53.1.83.

Schechter, J. (2010). The reliability challenge and the epistemology of logic. *Philosophical Perspectives, 24*(1), 437–464. https://doi.org/10.1111/j.1520-8583.2010.00199.x.

Schechter, J. (2013). Could evolution explain our reliability about logic? In *Oxford Studies in Epistemology, 4*, 214–39. https://doi.org/10.1093/acprof:oso/9780199672707.003.0008.

Schechter, J. (2017). Explanatory challenges in metaethics. In D. Plunkett & T. McPherson (Eds.), *Routledge handbook of metaethics*. Abingdon: Routledge.

Schechter, J. (2018). Is there a Reliabilit challenge for logic? *Philosophical Issues, 28*, 325–347. https://doi.org/10.1111/phis.12128 September.

Schlesinger, G. N. (1987). Miracles and probabilities. *Noûs, 21*(2), 219. https://doi.org/10.2307/2214915.
Schlesinger, G. N. (1991). *The sweep of probability*. Notre Dame: University of Notre Dame Press.
Street, S. (2006). A Darwinian dilemma for realist theories of value. *Philosophical Studies, 127*(1), 109–166. https://doi.org/10.1007/s11098-005-1726-6 Springer.
Urbach, P. (1992). Review of: the sweep of probability, by George N. Schlesinger. South Bend, IN: Notre Dame, 1991. *Mind, 101*(402), 395–397.
van Inwagen, P. (1993). *Metaphysics* (1st ed.). Boulder: Westview Press.
van Inwagen, P. (2015). *Metaphysics* (4th ed.). Boulder: Westview Press.
White, R. (2000). Fine-tuning and multiple universes. *Noûs, 34*(2), 260–276. https://doi.org/10.1111/0029-4624.00210.
White, R. (2005). Explanation as a guide to induction. *Philosophers' Imprint, 5*(2), 1–29.
White, R. (2007). Does origins of life research rest on a mistake? *Noûs, 41*(3), 453–477. https://doi.org/10.1111/j.1468-0068.2007.00655.x.
White, R. (2015). The argument from cosmological fine-tuning. In G. Rosen, A. Byrne, J. Cohen, & S. V. Shiffrin (Eds.), *The Norton introduction to philosophy* (pp. 43–50). New York: Norton.