A PUZZLE ABOUT INFORMATION, PROBABILITY AND SURPRISE

by

Selorm Yao Ohene

A Thesis Submitted in

Partial Fulfillment of the

Requirements for the Degree of

Master of Arts

in Philosophy

at

The University of Wisconsin-Milwaukee

May 2021

ABSTRACT

A PUZZLE ABOUT INFORMATION, PROBABILITY AND SURPRISE

by

Selorm Yao Ohene

The University of Wisconsin-Milwaukee, 2021
Under the Supervision of Professor Michael Liston

Suppose 92 coins, flipped in succession, all come up heads. If we were previously confident that the process was fair, we would be *surprised* at this outcome. What, if anything, explains our surprise? And is it warranted? In what follows I do two things. First, I propose and defend an explanation of our surprise: we are surprised at the 92-head sequence, and various other sequences, because they are *patterned*. Second, Martin Smith (2017) has argued that, on an initial assumption that the coin-flipping process is fair, an observation of 92 heads does not warrant surprise. Against Smith, I argue that even if we *knew* beforehand that the coin-flipping process is fair, our knowledge is *defeated* by an observation of 92 heads. Under reasonable constraints on the prior probability that our initial assumption of fairness was wrong, an observation of 92 heads (or various other patterned outcomes) makes us practically certain that the process is unfair. As such, an observation of 92 heads does warrant surprise.

To Nana, Sammy, Senyo, Sena, Doris, Mercy, and Eseli

TABLE OF CONTENTS

LIST OF FIGURES

1 Introduction

This thesis explores the following puzzle. Suppose 92 coins, flipped in succession, all come up heads.[1] Let's stipulate that before the coins were flipped, we were confident that the process was fair: the coins weren't trick coins, the flips were independent, there were no hidden magnets, or anything. In fact, for our purposes, we might even suppose that the process *was* indeed fair in this way, and that we *knew* this beforehand. In such circumstances, we would likely be *surprised* at an outcome of 92 heads. This raises two questions. First, *why* would we be surprised? More precisely, is there an informative explanation of our surprise — one that, for instance, would be nontrivial and non-circular, and would also explain our surprise at certain other outcomes (say, the sequence HTHTHTHT… of alternating heads and tails), as well as our *lack* of surprise at other outcomes (say, some apparently arbitrary sequence beginning HTTHTHHT…)? And second, considering outcomes at which we would be surprised, such as 92 heads, would our surprise be *warranted*?[2]

In "Why Throwing 92 Heads in a Row Is Not Surprising" (2017), Martin Smith argues that there is no good explanation for our surprise, and that "we *shouldn't* feel surprised, that we have *no reason* to feel surprised and, if we do feel surprised, then we're being irrational" (p. 2).[3] I think this is a mistake. In what follows I'll try to argue as much, by proposing an explanation for our surprise that shows it to be warranted.

[1] Two details are salient here: first, that the number of coin flips is 92; second, that 92 coins are flipped once each, rather than one coin 92 times each. I have left it open who is flipping the coins, or how they are being flipped, for two reasons. First, these are the terms in which Smith discusses the puzzle. (The initial case which provides the motivation for his account is more concrete, but the additional details do not feature prominently in his discussion.) Second, although the particular details of each case will affect the specific conclusions we are entitled to draw, I want to articulate a sufficiently context-general framework for thinking about cases of this type.

[2] I will use the terms "warranted" and "justified" (and their respective cognates) interchangeably.

[3] All undated page number references will be to this paper.

It will be useful to begin by briefly discussing some of the conceptual structure of surprise, focusing on its relationship to warrant. We'll do that in chapter 2 below. Then, in chapter 3, we'll examine two criteria Smith proposes to explain our surprise, and show why they don't work. In chapter 4 I'll give a suggestion of my own: we are surprised at highly *patterned* sequences. In chapter 5, I'll argue that the presence of pattern not only explains, but warrants our surprise: we are surprised at patterned sequences because they are unlikely to have been produced by 92 fair coin flips. Finally, in chapter 6, I will explore the prospects for extending the insights gained from this puzzle to a more general theory of surprise.

2 The nature of surprise

I'll try to clarify some of the conceptual structure of surprise in this chapter, focusing primarily on its relationship to warrant. Three points will be crucial: (1) responses of surprise are subject to *warrant*, (2) surprise itself serves as a *warrant for action*, and (3) whether or not surprise is warranted depends on the *initial* conditions in which it arises.

2.1 Warrant

Although all sorts of things might *elicit* surprise, not all of them *warrant* surprise. For instance, if you are surprised by something you should have been expecting, your surprise is unwarranted. Depending on context, expressions of the form "X is surprising" can mean either of two things: that X merely elicits surprise, or that X warrants surprise. In what follows, I will only say that X is *surprising* to convey the second sense, and I'll say X *elicits* surprise to convey the first sense.

2.2 Surprise as warrant for action

Besides the fact that circumstances serve as a warrant for surprise, Smith notes (and I agree) that surprise itself serves as a warrant for action (p. 6). If X surprises us, it strikes us as somehow abnormal and/or in conflict with our prior beliefs or expectations, driving us to inquire into an explanation of X and/or to revise those beliefs or expectations.[4] As such, for surprise to be

[4] In classifying belief revision among the "actions" warranted in the face of warranted surprise, I'm not endorsing doxastic voluntarism. Rather, I'm appealing to the idea that belief formation, maintenance and revision are *rational responses* to evidence, and hence expressions of *agency* or *rational activity* in a certain (we might say Kantian) sense, although this is, as Matthew Boyle puts it, "a notion of rational activity [...] broader than the notion of voluntary rational action" (2011, 144). We speak of people *changing* their minds, or *refusing* to do so. By contrast, falling is a mere causal effect of losing one's footing, certainly not a rational response, and not the sort of thing that can be *warranted*. Anyway, the main point is that belief revision is amenable to warrant. Whether or not one wants to call it an action, as I do, is secondary.

warranted, these actions must also be warranted, including whatever inquiry we might pursue into an explanation of X. Given this connection between surprise and inquiry, it's worth asking the following question: is the warrant for surprise settled at the *beginning* of inquiry, or does it depend partly on its *results*?

2.3 Warrant for surprise is determined in initial circumstances

Here Smith and I disagree. I claim the warrant for surprise is settled at the *beginning* of inquiry. In particular, surprise at X can be warranted if it is *initially* plausible to inquire into an explanation of X, even if such inquiry *concludes* that nothing was out of the ordinary after all. We can say, for instance: "Surprising event X occurred, but it turned out to be a coincidence." The fact that it turned out to be a coincidence upon inquiry doesn't make our initial surprise any less warranted.

Smith takes the opposite view. He suggests that if, upon investigation, the process is found to have been fair, then there's nothing surprising about the outcome.[5] I simply note this difference here, but will return to it in chapter 5 below.

[5] In making this suggestion, he runs together whether or not an event initially seems to require (and ultimately has) an explanation and whether or not it's surprising. In general, explanation-worthiness and surprise aren't equivalent, but for the purposes of the coin-flipping scenario in this paper we can assume they are.

3 Two bad criteria

Having clarified some of the conceptual structure of surprise in the previous chapter, let's move on to determining what might explain our surprise. An initial difficulty is that two apparently plausible criteria for explaining our surprise don't in fact explain it. This makes it easy to believe that there *isn't* any plausible explanation for our surprise, and consequently that it can't be justified. We'll see in chapter 4 that there *are* plausible criteria, but let's get the misleading ones out of the way first.

3.1 Mere unlikelihood

Smith offers a first suggestion for why we might be surprised at 92 heads: "it's very *unlikely* for someone to throw 92 heads in a row. And if something very unlikely happens, then that's got to be surprising, [hasn't] it?" (p. 3) That is,

(C1) A sequence elicits surprise just in case it is unlikely.

As Smith rightly points out, however, the 92-head sequence is as unlikely an outcome of 92 fair coin flips as some sequence HTTHHTHT…, say. Either sequence, or any other, will appear with probability 1 in 2^{92}, or about 2 in 10^{28}. If this is what *explains* our surprise, argues Smith, then our surprise is unwarranted, because if this improbability warranted surprise, then *every* outcome would warrant surprise — an absurd conclusion.

Although one certainly wants to say that unlikelihood is a *necessary* condition for warranted surprise, I agree with Smith that the unlikelihood of any individual sequence, in this sense, is insufficient: it wouldn't warrant surprise all by itself. But unlikelihood in this sense doesn't in fact *explain* our surprise, anyway. If we are initially confident that the process is fair,

we won't in fact be surprised at every outcome.[6] So let's have a second attempt at explaining what elicits our surprise.

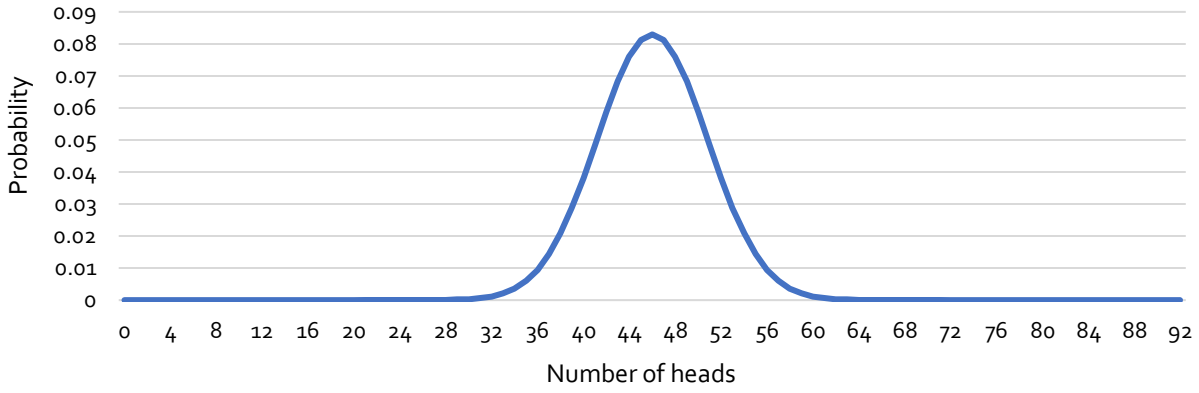
3.2 Zero subjective probability

Some basic probability theory will be useful to illustrate the second proposal. My exposition follows Smith's (pp. 4-5). Flipping 92 coins in succession is an instance of a *random process*, with various potential *outcomes* (here, sequences of heads and/or tails of length 92). The set of all possible outcomes (here, all such sequences) is the *sample space*. Now, typically, in thinking about the probability of obtaining 92 heads, we wouldn't just compare the 92-head sequence with other individual sequences. Rather, we'd divide up the sample space by number of heads and consider the probability of obtaining each possible value from 0 to 92 — in the language of probability theory, we'd take the distribution of the *random variable* [number of heads] on the sample space[7] — resulting in what's called a *binomial distribution*:

[6] To be sure, someone might still attempt to explain his surprise at 92 heads by saying something like, "What are the chances of that?" Now, if what he means is, "What are the chances of a fair process yielding 92 heads, *rather than some other outcome*?" then he is simply mistaken about what actually elicits his surprise, because he would not respond in the same way to all the other, equally unlikely outcomes. However — to foreshadow chapter 5 — he might instead mean something like, "What are the chances of a fair process, *rather than an unfair process*, yielding 92 heads?" And, under reasonable prior assumptions concerning the relative likelihood of various generating processes, he would be right, and moreover his surprise would be warranted precisely on that basis.

[7] For our purposes, a random variable is a function that assigns each outcome a unique *value*, such that the set of outcomes with a given value, or collection of values, has a well-defined probability. Here, the possible values of our random variable [number of heads] are the integers 0 through 92. A random variable yields a partition of the sample space, but note that multiple random variables can correspond to the same partition; we might then say the random variables in question are *equivalent*. For instance, the random variables [number of tails] is equivalent to [number of heads]: if you know the number of heads of a given sequence, you can immediately determine the number of tails, and vice versa.

Figure 1. Probability of obtaining a given number of heads



As the figure indicates, the distribution has a *peak* at the value 46, which has a *probability mass* of about 0.083. By contrast, the values 0 and 92 lie at the extreme *tails* of the distribution — the extreme ends of the graph, where it "tapers off"[8] — and have the lowest possible probability masses of about 2 in 10^{28}. Beyond about 63 heads, the probabilities are too low to register properly on the graph, but the figure consequently doesn't do justice to how sharply the probabilities drop as the number of heads increases. To illustrate: near the centre of the distribution, you're about one-ninth as likely to observe 56 heads as to observe 46. But out in the right tail, the probability of observing 86 heads is one in one billion the probability of observing 76. More generally, for any number of heads n greater than 46 and any additional number of heads k, the probability of observing $n + k$ heads becomes a much smaller fraction of the probability of observing n heads as n increases.

[8] The word "tail" Is liable to cause confusion: it may refer either to the side of a coin opposite heads, or to the extreme end of a distribution, where it tapers off. I'll try my best to make it clear which sense is at play, but both senses will be essential to my argument, so using it in both senses is unavoidable. For instance, as we'll see in chapter 4, an important feature of certain surprise-eliciting sequences isn't just that their values in some distribution are relatively unlikely, but that they're relatively unlikely *and* at the extreme ends of the distribution — that is, in the tails.

In short, although each *individual* sequence of coin flips is equally probable, when we divide the sample space up by number of heads, 92 is *far* less likely to appear than the other values. It may consequently be tempting only to entertain the possibility that values around 46 heads will appear — i.e., simply to *rule out* an outcome of 92 heads (pp. 4-5). And if we have ruled out the possibility, its appearance will elicit surprise. Hence Smith's second suggestion:

(C2) A sequence elicits surprise just in case we have ruled out its possibility prior to observing it.

Now, in the statistical sense, the *expected* number of heads (i.e., the probability-weighted average) is 46, and the bulk of the distribution is around 46. But just because some range of values is closer to the "expected value" (in the statistical sense), argues Smith, doesn't mean we should expect (in the everyday sense) to observe values inside that range.[9] As he puts it,

> there is one sense in which we should "expect" to get around 46 heads — we should regard this as highly likely, or assign it a high probability. The set of sequences in which we have around 46 heads covers a large proportion of the total set of outcomes. But there's another sense in which we shouldn't "expect" to get around 46 heads — we shouldn't *believe that this is going to happen*. We shouldn't believe that the sequences outside the set won't come up, while keeping an open mind about the sequences inside the set. There are no grounds for this — the sequences are all on a par. (5)

It is true that if we've ruled out an outcome, then its appearance would elicit surprise: in other words, ruling out an outcome is sufficient for surprise. Moreover, I agree with that we *shouldn't* rule out an outcome of 92 heads, if we assume that the process is fair. If this were what explained our surprise, it would indeed be unwarranted. But, again, I don't think this *does* explain our surprise. Smith identifies "expecting around 46 heads" (in the everyday sense) with ruling out

[9] Sometimes we *cannot* expect (in the everyday sense) to observe the expected value (in the probabilistic sense). The distribution of the number of children people in a large room happened to have might yield an expected value of 2.4, but it'd be patently absurd to expect that anyone have 2.4 children.

values significantly far from 46, but this identification is too strong.[10] If we are judging based on the distribution of number of heads, or an intuitive understanding of it, to say we "expect around 46 heads" isn't to say we *rule out* the possibility of 92 heads, assigning it zero probability of appearing. Rather, we deem it *highly unlikely* relative to the alternatives.[11]

So Smith's second proposal for explaining our surprise is no good, and gives us no headway concerning whether or not our surprise is warranted.

3.3 Modifying the second suggestion

But perhaps this proposal was on the right track, and simply needs modifying to correct the false assumption that we rule out certain sequences, as follows:

(C2') A sequence surprises us just in case it corresponds to a relatively unlikely value of some random variable on the sample space.

[10] Why might Smith have characterized the proposal in this way? Perhaps because, having dismissed the proposal that certain sequences are surprising in virtue of showing up with *low* probability, he thinks any other proposal must involve expecting certain sequences to show up with *zero* probability. But remember, Smith only dismissed the low-probability proposal because, given that every *individual* sequence is equally improbable, none is *relatively* improbable. When we consider sequences by number of heads, though, some values *are* less probable relative to the others. The *individual* sequences remain equally improbable, of course, but at this point in the argument we're explicitly setting aside individual (or absolute) probabilities in favour of probabilities with respect to the distribution of [number of heads].

[11] There's an interesting question here: given that 92 heads *could* show up, albeit with low probability, should we be able to say that we believe it won't show up? What's at issue here is the relationship between beliefs and credences. Some people identify belief that X with credence 1, and belief that not-X with credence 0. Others take it that belief and non-belief kick in at some threshold value of credence, while yet others advocate that beliefs are superfluous and credences are all we need. We won't settle the issue here, but it seems like there *is* a sense in which we are justified in believing that 92 heads won't happen: the probability is so low that for all practical purposes we can write the possibility off. For instance, the probability of getting 92 heads is far lower than that of two people, selecting a random grain of sand on the surface of the earth, will choose the same one. And if it seems reasonable to say outright that two people *won't* choose the same grain of sand, even though it's possible, it's not clear why we shouldn't be able to say outright that 92 heads won't show up.

Now, in response to C2′, Smith argues that if our surprise partly depends on a choice of random variable, it cannot be warranted (p. 5).[12] If we partition the sample space up by number of heads, then some values are less likely than others, and outcomes with relatively unlikely values might therefore seem to warrant surprise. But if the sequences are all individually "on a par" in their *individual* probability, why *should* we partition the sample space by number of heads and judge surprisingness on that basis? A different partition might well yield the opposite result.

For instance, take some class of outcomes C which makes up most of the total probability *and* includes the sequence with 92 heads. C might be, say, the set of all sequences which do *not* have equally many heads and tails. But it could also be a wildly gerrymandered collection which simply happened to contain most of the possible sequences, including the one with 92 heads, but left out some sequences with 46 heads. Having partitioned the sample space into outcomes in C and outcomes outside C, if we wanted to continue to defend our surprise, it looks like we'd have to say one of two things. We might maintain that the sequence with 92 heads is surprising, but that some more "arbitrary"-seeming sequence which falls in C is not. But then, it seems, we'd be admitting that whether or not a sequence is surprising doesn't depend on how we divide up the sample space. Or we might say that surprise *is* contingent on a partition of the sample space, and that in this case a sequence with equally many heads and tails *is* surprising, because it falls outside C, but the one with 92 heads is not — surely a disastrous result!

So, it seems, this candidate criterion also falls by the wayside.

[12] But I've tried to separate this (forceful) challenge from his (false) assumption that our surprise comes from having ruled out certain outcomes.

4 A better criterion: surprise as noticing pattern

We've been trying to figure out which sequences of coin flips elicit our surprise. So far, the best we've managed to do is find one criterion that seems relevant: some outcomes correspond to relatively unlikely values of some random variable — for instance, 92 is a relatively unlikely value of the random variable [number of heads]. But Smith presented the following challenge: whichever way you slice up the sample space, each *individual* sequence is still an equally likely outcome of 92 fair coin flips. So, even if a choice of random variable partly *explains* our surprise at some sequences and not others, no such choice could possibly *warrant* surprise at some outcomes and not others.

In this chapter, we'll identify a criterion that does escape Smith's challenge. Our ultimate result will be that we are surprised by sequences when we notice *pattern* in them. Then in chapter 5 we'll see why patterned sequences *warrant* surprise.

4.1 Evaluating random processes

To begin, it's worth asking: *can* we define a criterion for surprise at certain sequences without reference to some random variable on the sample space? The answer is "only with great difficulty." There's very little at all we can say about the sample space without reference to *some* random variable on it. Even Smith's decision to consider each outcome separately amounts to a choice of random variable on the sample space. That is to say, it isn't obvious that the only probabilities we should be concerned with are the probabilities of each individual sequence

resulting from a series of fair coin flips.[13] A more productive question would be: which random variables on the space should we consider, and why?

Unfortunately, this question has no general, precise answer. But to make some headway, it helps to reconsider the more abstract perspective we've been taking so far. Although, from a purely *mathematical* point of view, we can define all sorts of random variables on our sample space, the coin-flipping process is not itself an abstract mathematical object, a mere set of strings of length 92 consisting of Hs and Ts. Rather, it is a concrete process, involving various agents and physical conditions. Attending to the real intentional and physical features involved in the process helps us determine what random variables will be relevant.

In essence, we want random variables which track *meaningful* properties of the coin-flipping process and its outcomes, and which are *informative* for picking out those outcomes that elicit our surprise.[14] An arbitrarily gerrymandered partition of the sample space doesn't track any meaningful properties: it doesn't tell us anything about the sequences that fall into a given set in the partition, because all they have in common *just is* belonging to that set. Other random variables track meaningful properties, but aren't *informative* regarding which sequences elicit or

[13] In fact, there is an even more radical worry in the vicinity. Random variables are defined relative to the sample space, but even our *choice of sample space* (i.e of what *counts* as an individual outcome of the coin-flipping process) might have been different. For instance, instead of individuating outcomes solely by the resulting sequence of heads and tails, we might have individuated outcomes according to the resulting sequence of heads or tails, *and* which hand produced each flip, or which surface each coin landed on. Of course, this would make it much harder to assign a probability to the individual outcomes. But these other ways of individuating outcomes could certainly be relevant for explaining our surprise. For instance, we might discover that the coin only lands heads consistently when it lands on the table, which might lead us to hypothesise that the table is magnetic. (Thanks to Kevin Dorst for raising this worry about outcome individuation.)

[14] It may help some readers here to recall David Lewis' (1983) notion of a "natural property"; just as the property *green* "carves nature at the joints" better than *grue*, one might say the random variable [number of heads] carves up the sample space more meaningfully than some arbitrary gerrymandering of the space. However, I do not claim that some properties are *objectively* more meaningful, or informative, than others. Rather, I think meaningfulness varies with context. In the case of coin-flipping, a property such as the number of heads is relevant to salient physical properties of the process (such as whether or not the coin is biased). Other properties may not be as relevant in this context, but may become relevant in other contexts.

might warrant surprise. For instance, we might well divide the sample space into sequences with more than ten heads and sequences with ten or fewer: this is a meaningful random variable. But we aren't surprised at a sequence of 92 heads *because* it has more than ten heads (even though it does). And so, with regard to the question of surprise at 92 heads, this random variable isn't informative: pointing out that the class of sequences with more than ten heads contains most of the probability mass *and* includes the sequence with 92 heads is neither here nor there.[15]

In virtue of what properties, then, might we be surprised at a sequence of 92 heads? A few plausible candidates: we might be surprised that the sequence displays so many heads, or so many more of one face than the other, or such a long run of heads (or of one face), or so few (i.e. zero) switches between faces, etc. In each case three conditions hold:

(1) there is a property which admits of degrees,

(2) sequences possessing the property to an extreme degree are less common,

(3) the sequence of 92 heads is one of those sequences.

In probabilistic terms, if we consider the random variables (1) [number of heads], (2) [difference between number of heads and number of tails], (3) [longest run of heads], (4) [number of switches between faces], the 92-head sequence falls into a tail of each resulting distribution. This suggests the following, more general criterion:

(C3) A sequence elicits surprise relative to some degreed property just in case there is a random variable tracking the relevant property, and the sequence in question falls into a tail of the resulting distribution.

[15] Of course, we *would* be surprised at a sequence with ten heads or fewer. (That said, the binary random variable [ten or fewer heads/more than ten heads] is rather coarse-grained.

This criterion picks out quite a few sequences which indeed elicit surprise. Given an appropriate choice of random variable, this criterion predicts, not only that 92 heads will surprise us, but also that the following other sequences will: any sequences with a large imbalance of one face over the other; any sequence in which successive faces switch very frequently or very infrequently; any sequence in which a very long run of one face shows up, etc.[16] But it doesn't give us the full story, for the following reasons.

(1) It undergenerates surprise-eliciting sequences. For instance, we'd be quite surprised to notice that a sequence of apparently fair coin flips corresponded to the first 92 bits of π, but it's hard to think of a random variable that could possibly help explain why.

(2) Even if the criterion *were* extensionally adequate, we aren't surprised by certain outcomes *because* we take them to fall into a tail of such-and-such a distribution. We don't need to know probability theory to be surprised. In fact, for a given property, we might not even be aware that fewer sequences possess that property to an extreme degree; our judgments may simply happen *in fact* to line up with this criterion.

(3) It remains vulnerable to Smith's challenge, because it's relative to a choice of random variable. If you were unimpressed by my contention that appeal to some choice of random variable is well-nigh unavoidable, you might think this criterion might *explain* our surprise, but couldn't conceivably *warrant* surprise. So I should find the criterion less than ideal if my aim is ultimately to argue that our surprise is warranted. In any case, some sequences — for instance, 92 heads — seem to come out as less likely relative to

[16] Considering a wider range of random variables beyond [number of heads] helps us to pin down some surprise-eliciting sequences that we might otherwise have missed. The sequence HTHTHTHT…, for instance, has 46 heads, falling right into the peak of the distribution [number of heads], and yet it surprises us. But it *does* fall into a tail of the distribution of [number of switches between faces].

multiple choices of random variable, suggesting that there is some more general, intrinsic characterisation of those sequences in virtue of which this is the case.

And indeed, I want to suggest, there is such a characterisation.

4.2 Surprise as noticing patterns

What surprises us about the 92-head sequence, I want to propose in this section, is that it exhibits a high degree of *pattern*. What I mean by "pattern" lines up with our ordinary notion: we might say some sequence of coin flips displays a pattern of constantly (or disproportionately) landing heads, maybe, or of two heads regularly following a tail. But a more precise, technical definition will be useful in what follows: a sequence is patterned if it is *compressible*, if, without losing information, we can specify it in shorter terms than spelling out each individual flip ("92 heads" or "'H'*92" instead of "HHHHHHHH…", say).[17] Inversely, a sequence is random if there is no efficient way to specify it short of spelling it out. And there are grades of randomness in between; the shorter the minimum possible description length, the more patterned the sequence. The minimum possible description length of such a sequence, considered as a binary string, is also called the algorithmic complexity or Kolmogorov complexity of the string.

Hence, we have our fourth criterion:

(C4) A sequence elicits surprise just in case it exhibits a high degree of pattern.

[17] The various insights into characterising pattern employed in this paper, including this definition of pattern, are deliverances of the subfield of theoretical computer science known as algorithmic information theory, or AIT. For a paper which doubles as a friendly philosophical introduction to AIT and a demonstration of its considerable theoretical utility, see Daniel Dennett's classic "Real Patterns" (1994), from which readers may recognise the definition; see also *Every Thing Must Go: Metaphysics Naturalized* by James Ladyman and Don Ross, with David Spurrett and John Collier (2010), especially chapter 4. For a more traditional introduction see Cover and Thomas (1991), chapter 7, or Grünwald and Vitányi (2008). The standard reference is Li and Vitányi (2019).

It can be proven, for many random variables, that any sequence located in a tail of the probability distribution over one of those random variables has a short minimum description length. So this criterion encompasses our initial suggestion above.[18] But it also captures other surprise-eliciting sequences that our previous proposed criteria might not have captured easily: consider the sequence corresponding to the first 92 bits of π, say, or the sequence consisting of "HHTT" 23 times. Moreover, we can define the algorithmic complexity of a sequence independently of any random variable, and even independently of assigning probability values to individual sequences, so Smith's challenge to the previous criteria doesn't apply to this one. But it does come with a slight caveat: we have to *notice* the patterns in question. If we don't recognise the bits of π, for instance, the sequence might not surprise us, but *were* we to recognise the pattern, we should find it extraordinary that an ostensibly random sequence of coin flips happened to spell it out.

So that's my suggestion: what *elicits* our surprise at a given sequence is a noticeable pattern. But why might patterns *warrant* surprise? That is the subject of the next chapter.

[18] Here's a sketch of the idea: if, say, the sequence falls into the tail of a given random variable, we can "summarise away" the relevant part of the sequence, and include caveats when necessary. For instance, a sequence with 90 heads might be described as "all heads except 50 and 71."

5 Warranted surprise as alternative explanation

In this chapter I'll be undertaking two tasks. In the first section, I will propose and defend a criterion for *warranted* surprise. In the second section, I will argue that this condition is met, in the coin-flipping case, by patterned sequences. As far as the first task goes, here's the basic idea: when we encounter a surprising sequence, what we're evaluating is the likelihood that this sequence was produced by 92 fair coin flips, *rather than some other process*. And under reasonable assumptions concerning the prior probability that the process *was* fair, an observation of 92 heads makes us almost certain that the process was *not*, in fact, fair.

5.1 The hypothesis and how it works

The central idea we'll be pursuing in this section is nicely captured by Roger White:

> The crucial feature of surprising events seems to be that they challenge our assumptions about the circumstances in which they occurred. If at first we assume that the monkey is typing randomly, then her typing "nie348n sio 9q" does nothing to challenge this assumption. But when she types "I want a banana" we suspect that this was more than an accident. The difference is that in the second case there is some alternative but not wildly improbable hypothesis concerning the conditions in which the event took place, upon which it is much more probable. (White 2000, 270)[19]

We will use the case of 92 heads as an illustration. We'll have to wait until the next section to argue that, when 92 heads show up, there indeed exists some "alternative but not wildly improbable hypothesis" which substantially increases the probability of 92 heads. For now, let

[19] You may be surprised to learn that I pieced together this account before learning of the prior existence of a similar theory. If you were, the theory itself might explain why: I might have come across White's paper or something similar first, but somehow I didn't. (Thanks to Kevin Dorst for bringing it to my attention.)

us assume that this is the case. Then we can show why surprise would be warranted in such a case by the following Bayesian argument.[20]

Consider two alternative explanations for an outcome of 92 heads, at the highest level of generality; one, that we're faced with a sequence of 92 fair coin flips — call this hypothesis "fair"; two, that we aren't so faced — call that "unfair." Then, given 92 heads ("92"), what's the probability that we're faced with 92 fair coin flips? By Bayes' theorem,[21]

$$P(\text{fair} \mid 92) = \frac{P(\text{fair})\, P(92 \mid \text{fair})}{P(\text{fair})\, P(92 \mid \text{fair}) + P(\text{unfair})\, P(92 \mid \text{unfair})}.$$

Now, we have assumed throughout that we are confident, and perhaps even *know*, that the process is fair. As I see it, the right way to quantify that confidence is by assigning P(fair) a high prior probability, but that prior probability should *not* equal 1. Even if we took ourselves to *know* that the process were fair, I take it, we should admit some remote possibility that we might be wrong. So let's estimate the prior probability of "unfair" at, say, 1 in 1 billion, or 10^{-9}. (Again, this value is primarily chosen for convenience here — in the next section, we'll tackle the problem of determining priors in a more principled way.) Let's also suppose, for now at least, that on the unfair scenario, each coin has (say) a 99 percent chance of landing heads, instead of a 50 percent

[20] I owe the idea of applying Bayes' theorem here to Nassim Taleb (2020, 53). He employs it to argue that if you take some data to be modeled by a normal distribution, and you observe a "large deviation," you probably have the wrong distribution. The binomial distribution on a large sample space approximates the normal distribution, and so the argument carries over quite easily. Roger White, on the other hand, derives his argument in the quoted passage from Paul Horwich, who proposes a similar Bayesian account of surprise to that we're pursuing here.

[21] Here's a brief explanation for the reader who might find it helpful. We're assuming 92 heads could have come up in two ways: through 92 fair coin flips ("fair"), or otherwise ("unfair"). We want to know the probability that we're in the "fair" scenario, given that we have 92 heads. This is denoted P(fair | 92), which reads "the probability of 'fair,' given 92." The probability of being in the fair situation *and* having 92 heads come up is just the probability of being in the fair situation, times the probability that the fair situation yields 92 heads: i.e. P(fair) multiplied by P(92 | fair). Similarly, the probability of being in the unfair situation *and* having 92 heads come up is given by P (unfair) multiplied by P (92 | unfair). When you add up these two probabilities, you get the probability of having 92 heads come up by one or the other situation. And so, to estimate the probability that the situation was fair, given that we have 92 heads, we divide the probability of 92 heads *and* a fair situation by the probability of 92 heads *and* either a fair or unfair situation, i.e. the probability of 92 heads, however it came up.

chance; this yields a value of about 0.397 for P(92 | unfair). (We'll revisit these assumptions presently.)[22] And recall: the probability of obtaining all heads, given the fair coin hypothesis, is 1 in 2^{92}, or about 2 in 10^{28}. (I'm using powers of ten and approximations below for visual ease, but I'll carry out the actual calculation with the precise figures.) Then plugging in the numbers yields, to two decimal places:

$$P(\text{fair} \mid 92) = \frac{(1\text{-}10^{-9})(2\times10^{-28})}{(1\text{-}10^{-9})(2\times10^{-28}) + 10^{-9}(0.397)} = 5.09\times10^{-19}.$$

There isn't even a question – the fair coin flip hypothesis is out the window!

Now let's try some numbers more favourable to the fair coin flip hypothesis. Crank the prior probability that there's something weird going on down to 1 in 1 trillion, and suppose — why not? — that our "unfair" scenario is such that that each individual coin lands heads with probability a mere 0.7, so that the likelihood of 92 heads, conditional on "unfair," dwindles to a tame 6 in 10^{15}:

$$P(\text{fair} \mid 92) = \frac{(1\text{-}10^{-12})(2\times10^{-28})}{(1\text{-}10^{-12})(2\times10^{-28}) + 10^{-12}(6\times10^{-15})} = 0.03.$$

Closer, but still not close.

In fact, conditional on obtaining 92 heads, in order for the probability of "fair" to reach 0.5 — in order for it to be, pardon the pun, a "toss-up" whether or not we're dealing with 92 fair coin flips — the joint probability of an unfair situation and 92 heads would itself have to be of the order 10^{-28}. If it is substantially higher, the conditional probability of a fair situation, given 92 heads, is going to be very low. A surprising outcome such as 92 heads will take the agent from

[22] Incidentally, this is one place where a detail of our puzzle — that it involves 92 coins being flipped once each, instead of one coin 92 times — becomes relevant. The likelihood that one coin is biased so as to land heads with a given probability is *a priori* much higher than the likelihood that 92 coins are biased in such a way.

practical certainty that the process was fair (i.e., P(fair) very near 1) to practical certainty that it was unfair (i.e., P(fair | 92) very near zero). What the agent had reason to believe was almost certainly true, she now has reason to believe is almost certainly false. And this is what warrants her surprise. But now it is time to take a closer look at how the probabilities in question are determined, so as to justify our claim that 92 heads is indeed a surprising outcome.

5.2 Determining probabilities

Recall White's suggestion in the previous section regarding the conditions for surprise: there must be "some alternative but not wildly improbable hypothesis concerning the conditions in which the event took place, upon which it is much more probable." In the previous section, we used two examples to illustrate how the existence of such a hypothesis might justify surprise: when the event in question occurs, our credence in whatever circumstances we took to obtain goes from very near 1 to very near zero. Note, however: White does not give an account of when such an "alternative, but not wildly improbable" hypothesis exists, either for his typewriter case or more generally. All he says is that we are surprised in cases which involve such a hypothesis. Our task is to supplement White's account with some method for determining when such an alternative hypothesis exists — in particular, we must figure out what counts as a "wildly improbable" hypothesis in the first place, and how to tell which alternative hypotheses are wildly improbable and which are not.

My suggestion, at least in the coin-flipping case, is that such a hypothesis exists when there is pattern in a sequence. My argument for this proposal is essentially an appeal to Ockham's razor. Descriptions of a sequence can be associated with possible explanations of the

sequence, or possible hypotheses for the process that generated the sequence. The more highly patterned the sequence, the shorter the minimum possible description length; the simpler the corresponding explanation or hypothesis; and, by Ockham's razor, the higher the prior probability we should assign to it. Like the notion of minimum description length, this general line of argument has its origins in a branch of algorithmic information theory, specifically in the field of algorithmic probability pioneered by R. J. Solomonoff (see for instance Li and Vitányi 2019, chapter 5). However, because many of the details of the formal theory depend on assumptions which do not necessarily transfer directly to physical situations such as the coin-flipping case, the argument here will proceed at a much higher level of generality. I now lay out the argument in more detail with some examples.

As we have already seen, we can describe the 92-head sequence briefly: say, via the expression "'H'*92". Now we might associate such a description with a computer program which outputs the given sequence. For instance, consider a program which produces the 92-head sequence as follows: it starts by printing "H"; at each subsequent step, it prints another "H," and it stops after printing 92 symbols. The point is that, to describe such a program, we do not need to specify the entire sequence to the computer, or to give explicit instructions about which symbol to produce at, say, step 46. Analogously, in thinking of a concrete situation which would yield 92 heads in a row, we don't need a different explanation for the outcome of each coin flip; we may simply posit that the coin-tossing technique or mechanism is such as to cause the coin to land the same face on each flip. Then we would need only to determine the desired face (heads) and the desired number of flips (92), and such a process would yield a sequence of 92

heads as output. Thus, we can use the description "'H'*92" as shorthand for the suggestion that the 92-head sequence was produced by such a process.

Here is another example. We might specify the sequence HTHTHTHT... by the abbreviated description "'HT'*46." What sort of process might this correspond to? Again, thinking of the question in terms of a computer program might be illustrative. Such a process would start with an "H"; would print a "T" if the previous symbol was an "H,"; and then would repeat this process 46 times. Again, instead of specifying each individual flip, we only need to specify two such flips to be repeated a given number of times.

We notice the following parallel: both of the above sequences have short descriptions, and the processes which might generate those sequences are accordingly simple. Now, by contrast, let us stipulate a Kolmogorov random sequence beginning HTTTHHTHTHTT..., say, which cannot be described efficiently short of spelling it out. What sort of process would yield such a sequence? Unlike the above examples, there is no simple rule ("always land the same way," "always alternate," "repeat a particular sequence however many times," etc.) which would yield this sequence. Any process which yielded this sequence would have to be at least as complex as the sequence itself, in the following sense: if we were to write a computer program to output such a sequence, we would, at the very least, have to specify the outcome of each individual coin flip. (We could, of course, come up with a very complex formula which yielded the given sequence, but such a formula would be more complicated than the sequence itself.)

The above proposal for moving from description length to the characterisation of a process is schematic, but will do for now. But how do we get from that to determining a prior? We are guided by two principles here. First, as suggested above, the length of a description

corresponds to the complexity of the associated process. Second, the length of a given description is inversely related to the prior probability we assign to the given process. Consider, for instance, our random sequence beginning HTTTHHTHTHTT…, which (by stipulation) has no description shorter than the length of the sequence itself, i.e. 92. Suppose that we think it was produced by some non-random process. As we suggested above, such a process would require at least as much complexity to describe as the sequence itself. As such, we shouldn't give this alternative hypothesis a prior probability of more than 2^{-92}, or 1 divided by the total number of sequences, because there's no reason to assign a higher prior probability to the idea that the process was configured to produce *this* particular random sequence versus any other. In fact, the prior probability we assign it will end up being significantly *less* than 2^{-92}, or else there will not be "enough probability to go around" for all the conceivable hypotheses explaining the outcome. If we write "AH" for the alternative hypothesis, then, the product P(AH) P(92 | AH) will be at least an order of magnitude less than 10^{-28}, and so this alternative hypothesis will fail to disconfirm "fair" by an appreciable amount, giving the result that our non-patterned sequence is not surprising.[23]

We may rephrase the suggestion in the previous paragraph as follows. Write S(*n*) for the total number of sequences which admit a description of length *n*. We are taking description length as a measure of simplicity, and given equally simple processes, we have no reason (in general) to consider one of them more likely *a priori* than another. Thus, given a description of length *n* for a given sequence, the prior probability we assign the associated process should be at most 1 in S(*n*). But most sequences admit descriptions of many lengths. If we wanted to set

[23] I'm expanding here on (very brief) remarks made by Paul Horwich (1982, 103), from whom White draws (with slight modification) his theory of surprise.

the prior probabilities in a canonical way for any sequence, the natural thing to do would be to select the *minimum* description length, i.e. the Kolmogorov complexity. Note that the number of sequences with Kolmogorov complexity *n* would then be at most the number of possible descriptions in the language of length *n*, and so this approach would have us well on our way to finding a value for S(n), and therefore to finding a prior for a given process.

However, at least two problems arise at this point. First of all, descriptions of a sequence, and thus description lengths, are relative to a choice of language. It can be proven that, under reasonable constraints on the language in which strings and their descriptions are specified, the Kolmogorov complexity of a string is invariant under choice of language up to at most an additive constant independent of the length of the string.[24] But the constant in question will typically tend to be quite large, and so it will only provide a useful constraint on the complexity of very long strings; that is, the Kolmogorov complexity of a string is only invariant *asymptotically*. Even worse, Kolmogorov complexity turns out to be uncomputable in general: it is impossible to define a simple function that, when given a string, outputs its Kolmogorov complexity (even within a fixed language).[25]

These problems pose a challenge for using Kolmogorov complexity as the foundation for a general theory of surprise, and we will return to them in the conclusion. However, as far as our coin-flipping puzzle goes, we can afford to be relatively informal, and not worry about the precise numbers too much. This is because their relative orders of magnitude are more

[24] For this result, called the invariance theorem, see Li and Vitányi (2019), section 2.1.
[25] For the result that the Kolmogorov complexity is uncomputable, see Grünwald and Vitányi (2008), section 4.3, or theorem 2.3.2 in Li and Vitányi (2019, 127).

important to establish the point. The central idea is as follows. We begin with the following assumptions:

(1) The agent is initially highly confident that the process is fair. Thus P(unfair), the agent's prior probability that the process is unfair, is very low. However, it is multiple orders of magnitude higher than 10^{-28}.

(2) The probability of 92 heads if the process is unfair, P(92 | unfair), is also multiple orders of magnitude higher than 10^{-28}. In particular, both P(unfair) and P(92 | unfair) should have high enough probability that their product, P(unfair) P(92 | unfair), is at least an order of magnitude higher than 10^{-28}.

As we saw at the beginning of this section, establishing these prior probabilities with high precision is a difficult affair. However, given that a sequence of 92 heads is about as simple a sequence of coin flips as one could hope for, the heuristic argument for assigning higher priors to simpler processes was meant to show that we can reasonably take assumptions (1) and (2) on board. And given these assumptions, the Bayesian calculation yields a significant drop in probability from P(fair) to P(fair | 92), which justifies our surprise at an outcome of 92 heads.

Note that these assumptions are compatible with both P(unfair) and P(92 | unfair) being very low, as we saw in the example computations in section 5.1 above. In the second computation, for instance, P(unfair) was of the order 10^{-12} while P(92 | unfair) was of the order 10^{-15}. Consequently, the probability of 92 heads (from either a fair or unfair process) was of the order 10^{-27}, while the joint probability of 92 heads *and a fair process* was of order 10^{-28} (or about one-tenth as likely), resulting in a value of P(fair | 92) of order between 1 in 10 and 1 in 100.

With a few tweaks, we can extend this analysis to other highly patterned sequences as well, to explain why they are surprising. In those cases, the alternative hypotheses that present themselves will be somewhat different. (For instance, if a series of coin flips spells out the bits of π, the alternative hypothesis that presents itself will not be that the process is biased towards heads, but rather that the process involves some computation of π.)

To reiterate: what our surprise at patterns tracks, I'm claiming, isn't quite the likelihood of one versus another *outcome* of a given process, but rather the likelihood that one versus another *process* yielded a given outcome.[26] Our psychological unwillingness to suppose that patterns might have arisen by chance is misleading if it leads us to say they *cannot* so arise, but it correctly indicates that they are overwhelmingly *unlikely* so to have arisen, rather than by some other process. Apparently, in this instance at least, we're intuitive Bayesians![27]

Smith has largely been taking the generating process as given, whereas I've just argued that, faced with a sequence of 92 heads in particular, but any highly regular sequence in general — regular in the sense discussed in the previous section — we should suspend and investigate our initial assumptions about the process that generated the sequence. Now, Smith does say that we should investigate in this way, but he doesn't seem to think it means we should be surprised. Here's what he says:

> Having observed a run of 92 heads in a row, one should regard it as very *likely* that the coins are double-headed or weighted. But, once these realistic possibilities have been

[26] Our surprise *does* (partly) track the likelihood of a given outcome from a given process. In order for another process to plausibly have generated the outcome of 92 heads, the outcome had to be unlikely relative to the assumption that the process is unbiased.

[27] Incidentally, this shift in perspective — from holding constant a generating process and considering the possibility of various outcomes, to holding constant an outcome and considering the possibility of various generating processes — mirrors a formal duality between the "traditional" (and more familiar) information theory pioneered by Claude Shannon, on the one hand, and AIT, on the other. For discussion see Grünwald and Vitányi (2008), section 5, and Cover and Thomas (1991), chapter 7.

> ruled out, and we know they don't obtain, any remaining urge to find *some explanation* (no matter how farfetched) becomes self-defeating. (p. 7)

Smith does not explain why he thinks we should consider an alternative hypothesis "very likely," but one can reasonably assume he has something like the above Bayesian analysis in mind. What's interesting, though, is that he defends the claim that 92 heads shouldn't surprise us by arguing that there *needn't be* an explanation. And this is where my disagreement with him on the relationship of surprise to inquiry becomes salient. Shortly after the above passage Smith states that "it's rational to be surprised by an event if and only if that event requires investigation and explanation" (p.7). He agrees that an occurrence of 92 heads requires investigation, but he does not think it ultimately *requires* explanation. It may turn out simply to have happened by chance. As such, he thinks, it does not warrant surprise.

But I find this conclusion puzzling. I agree that there's nothing surprising about the outcome *anymore*, once we've investigated and concluded either that it had an explanation or that it happened by chance. But I see no reason why that should render our initial surprise unwarranted. Even if we initially *knew* that the process was fair, we *lost* that knowledge precisely in virtue of having observed 92 heads, and only regained it after investigation. And that perfectly well warrants our surprise. I agree with Smith's claim that "part of the purpose of surprise is to spur us into action" (6); that is, to cause us to seek an explanation, to *investigate*. But I disagree with his inference that our surprise is only warranted if there *must be* an explanation to be found. What makes our surprise warranted is rather the suspicion that there *may* be a *simple* explanation to be found.

6 Final considerations

My aim in this final chapter is to touch briefly on two things. In particular, I want to highlight some questions which we left unaddressed surrounding our initial puzzle, and to consider the prospects for extending our framework beyond our coin-flipping case, and ultimately towards a more general theory of surprise.

6.1 Additional questions surrounding our puzzle

In what follows I'll consider one assumption of our earlier treatment, and one question for further investigation. First, the assumption: we took it for granted that there was only one competing hypothesis in the vicinity (namely, that the coin-flipping process was biased towards heads). In principle, one wants to say, the coin-flipping process might have been unfair in all sorts of ways: it might have been biased towards heads (with any range of biases), biased towards tails, engineered in such a way as to alternate faces on each turn, or whatever. In principle, we would have to include all of these priors in the Bayesian calculation. But as it happens, the simplification does not affect the argument too much, for two reasons.

(1) The hypothesis that the process is biased towards heads is about as simple an alternative hypothesis as one could possibly hope for. As such, whatever prior probability we assign to other alternative hypotheses will, at best, be as negligible as the prior probability that the process is biased towards heads. Moreover, most of these already unlikely alternatives are also incredibly unlikely to produce an outcome of 92 heads. For instance, if "unfair*" is the hypothesis that each flip has a 99 percent chance of landing *tails*, then $P(92 \mid \text{unfair*}) = (0.01)^{-92}$, or 10^{-184}. In short, any alternative hypotheses will themselves

be so unlikely, or so unlikely to produce 92 heads, or both, that they will not affect the results of the calculation appreciably.

(2) In any case, it's dubious that we do, or even could, compute anything like precise probabilities in our judgments of surprise, whether consciously or unconsciously. Even if we did, for argument's sake, the precise probabilities in question would vary depending on which surprising sequence appeared, among other factors.[28] A more realistic approach to the reassessment of probabilities that takes place in an instance of 92 heads is that it is *retrospective*: only *after* observing a highly patterned outcome would the agent admit a (remote) prior probability that the process might have been unfair, and *then* compute the posterior probability.

That is to say, although it is useful to consider the "prior" probability of an alternative hypothesis, a more comprehensive treatment would have to determine to what extent this model accurately captures the actual cognitive process of surprise in real agents, and how useful the model is in light of this degree of accuracy or lack thereof.

Now consider the following modification of our puzzle. Before the coins are flipped, a sealed envelope is placed on a table containing a Kolmogorov random sequence, that is, a sequence which cannot be significantly compressed. The coins are then flipped, and the resulting sequence is noted. Then the envelope is opened and the sequence just written down is

[28] To mention just one other factor: consider an outcome in which, say, 90 out of 92 flips land heads. For the purposes of carrying out a Bayesian computation, we might classify this outcome as an instance of the event (i.e. collection of outcomes) "90 heads." However, since an agent in such a situation would also be surprised, and would have similar suspicions about the nature of the process, had 89 coins landed heads, or 91, or some other very high number, we might also classify it under the event "85 or more heads," or "between 89 and 91 heads," or something. This would present a significant complication for my argument *if* our actual surprise depended for its warrant on the precise outcome of a Bayesian calculation. But I don't think it does — the constraints are far more permissive, depending only on relative orders of magnitude.

compared with the contents of our envelope. They turn out to be a perfect match. We have a problem: according to our theory, patterned sequences are surprising, and algorithmically random sequences are not. But here, we have a surprising occurrence of a random sequence! Similarly, consider the case we had at the end of section 3.3: the sample space is divided into some large class of outcomes C and its complement. We went along with the suggestion that a sequence should not be considered surprising simply depending on whether or not it lies in C, since C might be an arbitrarily gerrymandered set which illuminates nothing about the actual features (physical, agential or otherwise) of the concrete coin-flipping process. However, in light of our envelope case, that might have been too hasty. For, if C is large enough and its complement small enough, it seems as though there *could* be genuine grounds for surprise if sequences outside C consistently showed up, regardless of their intrinsic randomness or pattern. Perhaps (depending on the details of how C is chosen), there might be some hidden reason why a disproportionately small class of outcomes keeps showing up.

What should we make of such cases? One way of approaching them (although there might be others) might be to consider the notion of *conditional Kolmogorov complexity*, which is (roughly) the shortest input necessary to add to an already given input to yield a specified output. So, although the sequence is Kolmogorov random, it has low complexity *conditional on* the sequence written in the envelope (to which it is identical). In our initial case, no such mechanism was remotely salient, and so there was no reason to condition on any sequence or set of sequences. But the notion of conditional complexity may be useful for instances where non-random mechanisms are already salient beforehand.

6.2 Extending the framework

What might be involved in extending the framework we have explored to a more general theory of surprise? Our discussion concludes with two suggestions.

First, such a general theory would have to be applicable to a wide range of real-world processes. Coin flips are easily considered as binary strings; however, more complex processes would presumably need to be described in higher-level languages, or might even resist being translated fully into algorithmic terms. Despite the use we have found for such terms in considering our puzzle, it is less clear how these more complex processes might be treated in terms of algorithmic complexity. Second, a general theory would need to be more precise about the probabilistic terms we used in explaining when surprise is warranted. For instance, when a surprising event occurs, what precise prior (or range of priors) should be assigned to the circumstances assumed to obtain by default, the event in question, or the alternative explanation? What does it mean for such an event to be "much more probable" given the alternative explanation, or for the alternative explanation to be "not wildly improbable"? Might answers to each of these questions change depending on answers to the others? We did not fully work out these details, but they would be relevant for broadening the scope of the theory.

Although these questions are yet to be answered, the aim of this thesis has been to suggest how the notions of algorithmic information and probability that gave rise to them might usefully address at least one philosophical question about surprise, and perhaps a greater range of questions besides.[29]

[29] Many thanks to Dan Baras, Selim Berker, Kevin Dorst, Blain Neufeld, Yasha Sapir, Joshua Spencer, and the members of the UWM philosophy graduate writing workshop for much helpful feedback. Special thanks to Michael Liston, my advisor, and William Bristow and Nataliya Palatnik, my committee members.

REFERENCES

Boyle, Matthew (2011). "Active Belief." *Canadian Journal of Philosophy Supplementary Volume* 35: 119-147.

Cover, Thomas A, and Joy Thomas (1991). *Elements of Information Theory.* New York: Wiley Interscience.

Dennett, Daniel (1991). "Real Patterns. *The Journal of Philosophy* 88:1, 27-51.

Grünwald, Peter, and Paul Vitányi (2008). "Algorithmic Information Theory." In *Philosophy of Information*, eds. Pieter Adriaans and Johan van Benthem. Amsterdam: North Holland.

Horwich, Paul (1982). *Probability and Evidence.* Cambridge, UK: Cambridge University Press.

Ladyman, James, and Don Ross, with David Spurrett and John Collier (2007). *Every Thing Must Go: Metaphysics Naturalised.* Oxford: Oxford University Press.

Lewis, David (1983). "New Work for a Theory of Universals." *Australasian Journal of Philosophy* 61.4, 343-377.

Li, Ming and Paul Vitányi (2019). *An Introduction to Kolmogorov Complexity and its Applications,* fourth edition. New York: Springer.

Smith, Martin (2017). "Why Throwing 92 Heads in a Row is Not Surprising." *Philosophers' Imprint* 17:21, 1-8.

Taleb, Nassim (2020). *The Statistical Consequences of Fat Tails: Technical Incerto, Vol. 1.* STEM Academic Press.

White, Roger (2000). "Fine-Tuning and Multiple Universes." *Noûs* 34:2, 260-276.