Explanation and determination
Gijsbers, V.A.

Chapter 8

Indeterministic Explanation II

8.1 Introduction

In this chapter, we will be looking at indeterministic explanations of deterministically produced events. (A deterministically produced event is an event that is the outcome of a deterministic process. An indeterministic explanation is an explanation where the explanans does not deductively imply the explanandum, but implies a high probability of the explanandum.) The question is not whether such explanations *exist*: it seems clear that they do, and that they are in fact widely used. We know that birth-control pills do not work in all cases; we also strongly suspect that there is an underlying deterministic mechanism that is responsible for their working in some cases and not in others. Nevertheless, when asked "Why did Mary not get pregnant?", we may answer "Because she used birth-control pills, and the probability of getting pregnant when you use birth control pills is very low." This is an indeterministic explanation for what we believe to be a deterministically produced event.

What we want to know is *how good* these explanations are, and specifically, whether it is possible that the best explanation of a deterministically produced event E is indeterministic. Presumably, most philosophers would say that if we were to find out exactly why the pills did work for Mary (even though they did not work for some others), then the deterministic explanation which incorporated these details would be more explanatory, would give us more total understanding of the situation, than would the indeterministic explanation. That is, most philosophers would say that indeterministic explanations of determined events are, *qua* how much they explain, second-rate explanations, although we can have excellent pragmatic reasons to prefer them.

However, Michael Strevens has argued in his recent book *Depth: An Account of Scientific Explanation* [132] that in some circumstances indeterministic explanations are strictly *better* than deterministic ones; that we should prefer them even if a deterministic explanation is available. "Often the objectively best explanation of a phenomenon is a nonsimple probabilistic explanation. In particular, a complex probabilistic explanation is often better than a deterministic explanation, regardless of the epistemic, practical, and other circumstances." (Strevens 2008 [132], p. 365.) In this chapter I will consider his claims and argue that, to the contrary, deterministic explanations of determined events should always be preferred.

8.2   Preliminary argument: Galápagos finches

Before we start our discussion of Strevens's main argument – which will turn out to be a long and often technical journey into the realm of microconstant processes and macroperiodic distributions of initial conditions – I wish to present and criticise a much simpler argument that Strevens gives in support of his thesis that indeterministic explanations are often to be preferred to deterministic ones. This argument consists in pointing out that scientists often *do* prefer them because they fit better into the most celebrated scientific theories. Let us consider Strevens's example.

Why was a drought on Daphne Major (one of the Galápagos islands) in 1976 followed by an increase of the average beak size of finches on that island? The explanation given by Grant 1986 [35] is that during a drought the only food left on the island is large, tough-to-crack nuts. Finches with large beaks have a much easier time cracking these nuts, which means that they have a greater chance of survival. A standard evolutionary argument allows us to conclude that the average beak size will increase in the next generation. This completes the explanation. It is obviously indeterministic. All the large-beaked finches *could* have died, either through starvation or through other causes. Greater fitness does not guarantee a higher survival rate, it merely makes it more probable.

The important question for our purposes is whether it would be better or worse to give a deterministic explanation of the same event, an explanation that in this case would detail all the events that determined the survival of each of the individual finches, and the production and survival of their offspring. Strevens claims that such an explanation would be worse:

> A fitter variant's replacing a less fit variant is to be explained,
> then, by citing the factors in virtue of which the strike ratios for

survival and reproduction are higher for the victorious than for
the vanquished variant. Any more detail degrades the explana-
tion. Darwinian explanation is not probabilistic because of our
laziness or ignorance [...] [but] will always be, and should always
be, probabilistic. ([132], p. 390)

In this example the explanandum *E* is: "The average beak size of finches
on Daphne Major increased after the 1976 drought". This fact can be ex-
plained both by an indeterministic explanation involving fitness, and by a
deterministic explanation that follows the actual adventures of all the birds
on Daphne Major in 1976. Is it true, as Strevens claims, that the indeter-
ministic explanation (which I will call IE) is *better* than the deterministic
one (which I will call DE)?

Strevens would certainly have been right had he claimed that IE can be
published in an academic journal, while DE would never be accepted; or that
IE will make it into standard biology text books, whereas DE will not. But
in these cases, IE is preferred not because it better explains *E*. IE elucidates
and supports the theory of Darwinian selection, and can be easily generalised
to yield predictions about what will happen to the finches on Daphne Major
when the next drought strikes. DE, on the other hand, does not contain the
central terms of the Darwinian theory (such as fitness), does not (necessarily)
allow us to predict what will happen in future droughts, and does not serve
to elucidate and support Darwin's theory. These differences are certainly
enough to explain why biologists are more interested in IE than in DE, and
why they might not even consider DE part of biological science. After all,
specific biological events are of interest to scientists *only* when they can be
put within a theoretical framework and related to other biological event.

But – and this is crucial – elucidating Darwin's theory, predicting what
will happen in future droughts, interrelating distant biological events, and
so on, although activities of great scientific value and importance, are not
part of explaining *E*, the specific fact that the average beak size of finches
on Daphne Major increased after the 1976 drought.

It is all too easy to be misled by our preference for general explanations.
Especially in science, we are rarely interested in particular facts. Biologists
try to understand the change in beak size on a particular island in a partic-
ular year *not* because this is an event that they find inherently interesting,
but because they hope to generalise that understanding into a broader un-
derstanding of the processes that have shaped and continue to shape life.
After all, what is the biologist trying to understand when he (or she) investi-
gates the Galápagos finches? He is trying to understand the mechanisms of
Darwinian evolution; and in order to understand these, he creates and tests

an explanation that mentions these mechanisms. Given that this is what
interests him, it is natural that he seeks to gain only *that* understanding of
the Galápagos finches which *can* be generalised. Therefore, the biologist will
vastly prefer IE over DE, quite apart from the question which of the two
gives more understanding of the particular fact under consideration.

Someone who is not interested in Darwinian selection as such, but really
wants to understand the particular fact that the average beak size of finches
on Daphne Major increased between 1976 and 1977, would be interested in
more aspects of the situation than the biologist. For this person, it is relevant
to know that if the large-beaked finches 3, 16, 35 and 121 had not escaped
from nearly fatal encounters with predators, the average beak size would
have decreased rather than increased. This fact, although uninteresting to
the biologist, is nevertheless a bona fide difference maker. Knowing about this
difference maker increases my understanding of why the beak size increased
on this particular island in this particular year. DE gives us these facts;
IE does not. Therefore, DE is the better explanation of the explanandum in
Strevens's example. (The presuppositions of this argument will be considered
at length in the rest of the chapter.)

An interesting corollary of this discussion is that science rarely aims to
find the best explanations of individual events; it aims to find the best expla-
nations of general patterns (or general patterns of explanation, the difference
between the two being often only metaphysical). It then evaluates particu-
lar explanations on the basis of whether they can be generalised to explain
general patterns (or become general patterns of explanation). In so far as
the patterns are indeterministic (and they will often be, perhaps even always
in the special sciences), science will prefer indeterministic explanations. So
Strevens is right when he claims that "Darwinian explanation ... will always
be, and should always be, probabilistic"; but he is right not because indi-
vidual biological events are best explained indeterministically, but because
an explanation is not Darwinian unless it uses probabilistic terms such as
'fitness'.

8.3   Strevens's theory of explanation

We have seen that Strevens's thesis cannot be established by showing that
scientists prefer indeterministic explanations. After all, the scientist is often
more interested in constructing and testing general theories than in explain-
ing certain particular facts. But the empirical argument is not Strevens's
most important one. Before we go to the main argument, though, it will be
useful to give a quick summary of his 'kairetic' theory of explanation. We

will first talk about deterministic explanations.

> An attempt at a deterministic kairetic explanation of an event
> e begins with a veridical, deterministic atomic causal model for
> e. That the model is deterministic and causal means that (a)
> the setup of the model entails the target e, and (b) this entail-
> ment mirrors a real-world relation of causal production. Such
> a model is subjected to the optimizing procedure; the result is
> an explanatory kernel, a model containing only factors relevant
> to the explanandum. A standalone explanation is built from ex-
> planatory kernels. (Strevens 2008 [132], p. 358)

Explaining e starts by constructing a veridical, atomistic, causal model.
An atomistic model consists of (propositions affirming the existence of) two
or more states of affairs or events, one of which is the explanandum e, and
at least one generalisation, where this generalisation and the non-e states of
affairs together deductively entail e. In addition, the model must be veridical
(which means that all statements occurring in it are true), and the entailment
must be causal.

To define causal entailment, Strevens takes the relation of 'causal influ-
ence' as primitive. Causal influence is the relation that holds between any
cause and its effects, and it is specifically *not* a relation of causal relevance.
The planet Mars has some gravitational effect (albeit a very small one) on
a ball thrown on earth; this effect varies with Mars's exact position; and
therefore the exact position of Mars stands in the relation of causal influence
to the shattering of a window hit by the ball. By contrast, the exact position
of Mars makes no difference to the shattering of the window (had Mars been
at any other position near its actual one, the ball would still have shattered
the window), and is therefore not causally relevant to the shattering.

A causal entailment is an entailment every step of which "corresponds" to
an actual relation of causal influence, which means that the (states of affairs
whose existence is affirmed in the) premises are causal influences on the
conclusion by virtue of the (law given in the) generalisation. An entailment
from the throwing of the ball, the position of Mars, and Newton's laws, to
a shattering of the window is a causal entailment. An entailment in the
other direction would not be a causal entailment, since the shattering of the
window is not a causal influence on the throwing of the ball. (Strevens treats
the topic in much more detail: Strevens 2008 [132], pp. 74-83.)

This first step leaves us with a model that causally entails the explanan-
dum, but which may contain an immense amount of irrelevant information
(like the position of Mars, or the fact that the ball was orange, neither of

which makes a difference to the shattering). We get rid of these irrelevancies in the *optimising procedure*, where everything that is irrelevant to the actual obtaining of e is removed. Each premise that isn't necessary for deriving the explanandum must be removed and all the premises must be made as abstract as possible – as abstract, that is, as is compatible with the model remaining veridical, causal and deterministic. Abstraction here means loss of (logical) content: a proposition p is more abstract than a proposition q just in case q implies p but not vice versa. The idea is that everything that remains in the model after removal and abstraction is a real difference maker for, and therefore an explainer of, the explanandum.

In our example the gravitational influence of Mars and the colour of the ball would be removed from the model, since they are irrelevant to the breaking of the window. (In order to maintain entailment, it might be necessary to add a single, very abstract proposition saying that all unmentioned influences are negligible.) In addition, the optimising procedure will abstract away from the precise mass and speed of the ball, leaving only the proposition that the ball's momentum is greater than some threshold value needed to shatter the window. (The optimising procedure, which does most of the work in Strevens's account of explanation, faces grave problems – see Gijsbers 2009 [32]. However, these problems are not relevant for this chapter.)

When we have optimised a veridical, atomistic, causal model of e, we have what Strevens calls an *explanatory kernel* for e. Such a kernel is an explanation of e; and more complex explanations of e can be constructed by stringing such kernels together.

The *indeterministic* case is almost equivalent. The major difference is that the premises of a probabilistic causal entailment do not entail the conclusion, but only entail a certain probability for the conclusion. All things being equal, the higher this probability, the better the explanation, with the deterministic case being the optimum. However, Strevens will argue that all things are *not* equal, and that under certain circumstances, a loss in probability is more than compensated for by an increase in relevance. We will see this argument unfold in the next section.

8.4   Microconstancy and macroperiodicity

According to Strevens, there are three types of explanation of deterministically produced events: low-level deterministic explanations, high-level deterministic explanations and indeterministic explanations. (Not all events can be explained using all three types.) In order to discuss these types with clarity, he introduces the concepts of microconstancy and macroperiodicity, and

elucidates these using the example of a wheel of fortune (Strevens 2008 [132], p. 370 ff; see also Strevens 2003 [129], 2005 [131].). We will follow Strevens.

Imagine a wheel of fortune that has alternating red and black sections, and assume that the section in which the pointer will come to rest is a function of nothing but the wheel's initial spin speed v (the wheel is reset to the same starting position after every spin). Assume furthermore that there is a probability distribution over v that gives the probability that a spin will have a speed between any two values of v. Strevens then argues that if the wheel is 'microconstant' with strike ratio p, and if the probability distribution is 'macroperiodic', it logically follows that there is a high probability that the long-run frequency of red outcomes will be approximately p after a large number of trials.

What are microconstancy and macroperiodicity? Strevens most rigorously defines these notions in Strevens 2003 [129]. Since one of the main questions in what follows will be when macroperiodicity is well-defined, it is crucial to follow this more rigorous discussion. I will shorten and rephrase it; the curious reader should check out pp. 127-132 of Strevens 2003 [129].

Microconstancy: Let S be a function from a metric space I of initial conditions to a space of two possible outcomes (which we will call 'red' and 'black'). We will call all points in I which are mapped onto the outcome 'red' 'red points'; and all points in I which are mapped onto the outcome 'black' 'black points'.

S is *microconstant* with respect to a partition $\mathfrak{U}$ of I into measurable subsets of non-zero measure if:

1. Every set U in $\mathfrak{U}$ is contiguous.
2. Every set U in $\mathfrak{U}$ has the same strike ratio, where the strike ratio is the measure of the set that contains all red points in U divided by the total measure of U.[1]

Informally, a function S is microconstant just in case we can carve up the space of initial conditions such that each part has the same ratio of red to black points. In the case of the wheel of fortune, we may expect that as v increases, we will have alternating red and black sections which are the same size as their neighbours. Here S (the map from v to the outcome space {red, black}) would be microconstant, since we can cut up I into pieces that

---

[1] I have written "if" instead of "if and only if" because our current definition is somewhat too restrictive to capture Strevens's pre-mathematical idea. For the purposes of this chapter, however, we can continue as if the conditions we have given are necessary as well as sufficient.

comprise one neighbouring red and black section. For each such piece of I, the strike ratio would be 0.5: red is exactly as likely as black.

The reader may notice that the definition of microconstancy we have just given is trivial (as long as I is contiguous and measurable, and the sets of red and black points are measurable as well). We can always cut up I into exactly one piece, namely I; and in that case all pieces of I will of course have the same strike ratio. This problem will be solved when we define macroperiodicity, so without further ado:

Macroperiodicity: Let P be a probability distribution on a metric space I of initial conditions. Then P is (approximately) *macroperiodic* with respect to a microconstant function S on I if (and only if) there is a partition $\mathfrak{U}$ of I such that:

    1. S is microconstant with respect to $\mathfrak{U}$.

    2. For all U in $\mathfrak{U}$, P is (approximately) constant over U.

Informally, a probability distribution is macroperiodic with respect to a microconstant function just in case it changes slowly compared to the size of red and black regions. A macroperiodic distribution does not have significant peaks or dips that lie solely within a red or a black region. This means that the relative probability of two neighbouring regions will be roughly equal to their relative size: in whatever general area of I we are, the chance of getting a red outcome is p.

I will not prove it here, but if the probability distribution over the initial conditions of an experiment is macroperiodic with respect to the outcome function, we can deduce that the probability of getting red is equal to the strike ratio. We will shortly see that Strevens also claims that we can *explain* the probability by citing the microconstancy and the macroperiodicity.

In most realistic systems, we will have approximate rather than full macroperiodicity. How this is to be defined exactly will not concern us here; throughout this chapter, we will make no difference between full and approximate macroperiodicity.

8.5   Three types of explanation

We now return to Strevens's example of the wheel of fortune. We will assume that the probability distribution over the initial spin speeds is macroperiodic, and we will also assume that the laws that determine the motion of the wheel and the paint scheme are such that the map from the space of initial spin speeds to the space of outcomes is microconstant with strike ratio $p = 0.5$.

Imagine that the wheel is spun 500 times. Let E be the event that, of these 500 spins, between 200 and 300 end up pointing to a red section of the wheel. E is of course an extremely likely outcome of the experiment, and we will suppose that it actually takes place. What explanations of E can we give in this situation? Why does the number of red outcomes lie between 200 and 300? According to Strevens, there are three general types of explanation that can be given:

1. A **low-level deterministic** explanation. We detail the initial spin speed of all of the trials with just the precision necessary to entail in which section the pointer will end. We also give a precise description of the causal mechanisms which determine how quickly the wheel slows down, and we describe where the boundaries of the red and black sections are. Together, this information entails the outcome of each spin of the wheel; and therefore it entails E.

2. A **statistical deterministic** explanation. We first state the macroperiodicity of the *actual* initial velocities (the set of actually chosen initial velocities, not the probability distribution over them – we will say much more about this later), and we then state the physical properties by virtue of which the wheel of fortune is microconstant with $p = 0.5$. Together these facts entail E. We do not specify the causal mechanisms by which the wheel spins down, the actual physical places of the red and black boundaries, or any other details of the physical situations – nor do we need to in order to deduce and explain E.

3. An **indeterministic** explanation (which Strevens calls a "microconstant explanation"). We first state that the set of outcomes are the outcomes of a wheel of fortune. We then state that outcomes of a wheel of fortune have a high probability of being macroperiodically distributed. Finally, we state the physical properties by virtue of which the wheel of fortune is microconstant with $p = 0.5$. From this, we can deduce that with high probability, E.

According to Strevens, there is a hierarchy of explanatory goodness among these explanations: the indeterministic explanation is slightly better than the statistical deterministic one; and both are far superior to the low-level deterministic explanation. This, of course, is the claim that I want to examine and dispute.

We will first focus on Strevens's claim that the statistical deterministic explanation is better than the low-level deterministic explanation. It does in fact follow from his theory of explanation that this is the case. The first thing

we must see is that the premises of the statistical deterministic explanation are entailed by the premises of the low-level deterministic explanation. That the actual distribution of initial velocities is macroperiodic must be implied by the full characterisation of the initial velocities; and the physics of the wheel plus the paint scheme will entail that the wheel of fortune is microconstant with $p = 0.5$. Thus, the low-level deterministic explanation contains strictly more information than the statistical deterministic explanation; and yet they lead to the same conclusion, namely, E. But we have seen that according to Strevens, explanation involves an optimising procedure that takes away as much information as possible while leaving the entailment intact. Consequently, the low-level deterministic explanation must be defective: if it had been optimised thoroughly, its irrelevancies would have been removed, and the statistical deterministic explanation would have resulted.

My counterargument will be technically involved, but simple in intent: I wish to show that the statistical deterministic explanation does not exist. And not only is there no statistical deterministic explanation of E, there are no statistical deterministic explanations *at all*. If these do not exist, the argument that shows that low-level deterministic explanations are defective immediately disappears.

8.6   The statistical deterministic explanation does not exist

8.6.1   The problem

We must now notice a problem that besets the statistical deterministic and the indeterministic explanation. The statistical deterministic explanation has as one of its premises that the actual initial velocities are macroperiodically distributed. The indeterministic explanation has as one of its premises that the set of actual initial velocities of a wheel of fortune have a high probability of being macroperiodically distributed. But, and this is the problem, the set of actual initial velocities is not a probability distribution, and therefore cannot be macroperiodic in the way that Strevens and I have defined it.[2]

I will argue that this problem is fatal for the statistical deterministic explanation: there is no redefinition of macroperiodicity that will allow us to successfully apply the concept to sets of actual initial conditions. (This is not fatal for the indeterministic explanation, since it can easily be rephrased to

---

[2]Strevens's ideas of microconstancy and macroperiodicity have been discussed by other authors; see, for instance, Kronz 2005 [61] and Sklar 2006 [125]. These discussions do not bear on our current problem.

mention the macroperiodicity of the probability distribution P rather than the macroperiodicity of the set of actual initial conditions.) In order to show this, I wish to consider some *prima facie* promising ways of extending the definition of macroperiodicity to sets of outcomes.

What are the minimal requirements for a definition of the macroperiodicity of a set of actual initial conditions? At least the following three:

1. The definition must be strong enough to support the entailment in the statistical deterministic explanation. Macroperiodicity plus microconstancy must entail E, otherwise we do not have a *deterministic* explanation.

2. The definition must be such that the resulting entailment is an *explanation*. (We cannot have the conclusion as one of the premises, for instance.)

3. The definition must be fulfilled by all (or at least most) 'typical' sets of initial conditions generated by a macroperiodic experiment. Otherwise, we are not giving a definition of macroperiodicity.

Can we find such a definition?

8.6.2 Attempts to define macroperiodicity

It is perhaps most promising to start with the set of actual initial conditions, give an algorithm to generate a probability distribution from it, then check whether this probability distribution is macroperiodic. Let $\{O_1, O_2, \ldots, O_N\}$ be a set of N points in I. Let $P'(x)$ be some function that maps all points $x \in I$ to probability distributions over I that peak at x and monotonically decrease as the distance from x becomes greater. We can then create a probability distribution P from the set O through the following definition:

$$P(x) = \sum_{n=1}^{N} \frac{P'(O_n)}{N}. \tag{8.1}$$

Informally, what we have done is taken every actual initial condition and replaced it by a little probability bump. This means that in areas where many of the actual initial conditions were found, the probability in our distribution will be high; and in areas were few of the actual initial conditions were found, the probability in our distribution will be low. We now define that the set O is macroperiodic if and only if $P(x)$ is a macroperiodic probability distribution with respect to the outcome function S.

Is there a $P'(x)$ for which this definition meets both the requirements given above? For simplicity, we will discuss only the example of the wheel of fortune. Suppose first that $P'(x)$ is a wide distribution compared to the average distance from a red point to the nearest black point, or from a black point to the nearest red point. In that case, it is easy to get a macroperiodic distribution $P(x)$, since the value of $P(x)$ will tend not to vary much between neighbouring black and red areas. Indeed, it is *too* easy to get a macroperiodic distribution: even if all the actual initial conditions lie in red areas, we would still have a macroperiodic $P(x)$. Thus, there would be no entailment from macroperiodicity to getting the expected number of red outcomes.

Suppose then that $P'(x)$ is a narrow distribution compared to the average distance from a red point to the nearest black point. (We will look at distributions that are tailor-made to distribute probability only over the red or black section in which the point falls later.) Now the problem is that it is much too difficult to generate a macroperiodic distribution $P(x)$. Remember that a macroperiodic distribution does not vary appreciably over *any* member U of $\mathfrak{U}$. If our $P'(x)$ is narrow, this means that *each* neighbouring pair of red and black sections must contain points in the ratio $\frac{p}{1-p}$. This will almost never happen. It cannot happen if N is small compared to the number of sections, and we still do no expect it to happen if N is large (but not very large) compared to the number of sections.

This is a problem for three reasons. First, it would make the statistical deterministic explanation unavailable in many real-world situations. Second, it would fail to meet requirement 3 mentioned above. And third, there must be something wrong with a definition of macroperiodicity that can be instantiated only when the number of sections is small compared to the number of trials, because it is obvious from our initial discussion of macroperiodic probability distributions that macroperiodicity is *easier* to achieve when the number of sections is larger.[3] We must be on the wrong track entirely.

Instead of generating a probability distribution from O, we might want to check whether O is a probable distribution given P. (I will assume that we already know what P is; if we do not, even fewer strategies are open to us.) But if the number of points in I is large compared to the number of trials – and this will always be the case if our initial conditions are real-valued – almost all sets will be equally likely. This problem can be circumvented

---

[3]But isn't it the case that statistical explanations will work only when the number of trials is large compared to the number of outcomes? Yes, but outcomes and sections are not the same. In our example, whatever the number of sections, there are exactly two outcomes: "red" and "black". We can thus give a statistical explanation of the (expected) result of a set of five hundred wheel of fortune spins because five hundred is much bigger than two. The number of sections is irrelevant to this.

only by cutting I up into pieces, and counting the number of actual initial conditions that fall into each of the pieces. However, that is merely a variant of the previous method, with all the same problems associated with it.

A third possibility would be to check whether O is likely to have been generated from a macroperiodic distribution, that is, whether it is probable given O that the probability distribution that generated O is macroperiodic. The step from outcomes to the probability of the underlying distribution is not trivial, but might be made with a kind of Bayesian updating. Now the probability distributions that most increase in probability, given O, will *not* be macroperiodic probability distributions but distributions that are sharply peaked at the values found in O. So looking at the greatest increase in probability will never give a macroperiodic distribution. Alternatively, we may look at the most likely distribution. But either the macroperiodic distributions did not start out with a higher initial probability than the non-macroperiodic ones, in which case the greatest increase in probability equals the greatest final probability, and a non-macroperiodic distribution will always win; or the macroperiodic distributions did start out with an advantage, in which case small sets of trials will always come out as being macroperiodic, even when all of them lead to red outcomes.[4]

Looking at specific proposals has given us more insight in the difficulties that we encounter when we try to speak about a macroperiodic set of initial conditions. But we can also give a general argument against all these approaches, namely, that no distribution of points can imply that the expected ratio of black and red outcomes has been achieved unless detailed knowledge about S is added. For even if the points are perfectly distributed over I, and even if they are perfectly distributed within each member of $\mathfrak{U}$, there is nothing to stop them from all being red points or all being black points. S, after all, can be *any* distribution of black and red as long as these colours appear in the same ratio within each member of $\mathfrak{U}$. So it is impossible to get the entailment we need without using detailed knowledge of how S divides I into red and black regions. Let us look at definitions of macroperiodicity that use such detailed knowledge of S next.

8.6.3 Further attempts

No finite set of points in I, no matter how randomly distributed, is guaranteed to lead to the expected ratio of red to black outcomes. We must have detailed

---

[4]Yet another alternative is to check whether the total probability of macroperiodic distributions exceeds the total probability of non-macroperiodic distributions. Evaluating this proposal would be technically more involved, and is unnecessary given the more general arguments to which we will now proceed.

knowledge of S in order to define a property that *does* guarantee this ratio.

The most straightforward approach would be to say that a set of points is macroperiodically distributed if and only if for all $U \in \mathfrak{U}$, the ratio of red points to black points is approximately $\frac{p}{1-p}$, with p the strike ratio. But of course this does not work. That I can be cut up into many parts that all have the expected ratio of red to black points *implies* that the total ratio is the expected ratio, but it does not *explain* it. The explanation is not exactly circular, but it is utterly unenlightening. Hearing that there are more red cars than black cars in each European country does not make us understand why there are more red cars than black cars in Europe.

Can we give a general argument that shows that any definition of macroperiodicity that involves a precise knowledge of S is bound to be unexplanatory? I think we can.

Remember that Strevens's theory is a theory of causal explanation. Both he and I believe that any good explanation of the outcomes of a wheel of fortune must be a causal explanation, and in Strevens's terminology this means that the entailment that forms the core of the explanation must be a causal entailment. This more or less means that we deduce an effect from its causes.

Let us think of the indeterministic explanation for a moment. Here, we talk about the macroperiodicity of a probability function. In the case of a physical process like spinning a wheel of fortune, this macroperiodicity is a physical property of those things that spin the wheel – properties of me and my muscles and my brain, for instance. Stating that the probability function over the initial spin speeds is macroperiodic is stating a high-level physical fact about me and my interaction with the wheel of fortune. In the same way, the microconstancy of the wheel of fortune is a high-level physical fact about the wheel. These two high-level physical facts will with high probability cause a set of outcomes where red occurs in a fraction p of the cases. Thus, there is causal entailment.

But now return to the deterministic statistical explanation. Here the macroperiodicity of a set of initial conditions is not a high-level physical property of the process that produced them, just as the fact that a die came up 6 is not a physical property of the process of rolling a die. Instead, it is a functional property of this instantiation of the process, but Strevens and I agree (Strevens 2008 [132], pp. 386-388) that functional properties like these do not explain the outcomes of wheel spins or die tosses. The fact that a die toss had "initial conditions which led to the die coming up 6" does not explain why the die came up 6. Why not? Because here we have a purely logical, not a causal entailment. Strevens writes:

> Now, in order for the model to be a causal model, the entailment must be a causal entailment, which is to say that the properties figuring in the premises must play a part in the entailment that reflects their role in the causal production of the outcome. It is this condition that is not satisfied in the fully functionalized model. The course of the causal events is as follows: the spin speed, in virtue of its magnitude and various physical properties of the wheel, determines a final resting place for the wheel. The final resting place, in virtue of the position of the pointer and the wheel's paint scheme, determines the outcome, red or black. The paint scheme enters into the causal story only once the final resting place is determined, then, so it should not enter into the deduction of the outcome until the final resting place is deduced. (Strevens 2008 p. 387.)

Strevens is absolutely right: the paint scheme cannot enter into our explanation *before* the resting place of the wheel is determined. If it does, we can be sure that a causal entailment has been turned into a logical entailment through functional properties. But of course the paint scheme is S, and we have already shown that S *must* enter the definition of the macroperiodicity of a set of initial conditions – if it does not, we cannot get entailment of the explanandum.

On Strevens's own theory of explanation, then, the statistical deterministic explanation must fail either because it does not entail the explanandum (and is therefore not a deterministic explanation), or because it is a functional pseudo-explanation. In other words, statistical deterministic explanation is impossible. Such explanations do not exist.

8.7 The low-level deterministic explanation is not irrelevant

If the statistical deterministic explanation does not exist, it no longer follows that the low-level deterministic explanation contains irrelevant information. Perhaps *all* the information it contains is needed to causally entail the fact that approximately 500 of the 1000 trials with the wheel of fortune ended up red. Strevens does not believe this to be the case:

> [The low-level model] contains far more detail than is necessary to entail the frequency; it is therefore, unlike the [indeterministic] model, bloated with causal irrelevancies. For example,

whereas the low-level model must concern itself with the magnitudes of...the frictional forces acting on the wheel, the [indeterministic] model needs state just the physical properties of the wheel in virtue of which the forces have a circular symmetry.... The frictional details are therefore causally irrelevant to the frequency – only the physical basis of the frictional symmetry matters – yet the low-level model cites the details all the same. ([132], pp. 380-381)

Before we attempt to decide whether the low-level model contains irrelevancies, we must do two things. First, we must use Strevens's own theory of explanation to construct the best possible low-level explanation. Second, we must say what we mean by irrelevancies.

Remember that this was our initial characterisation of the low-level explanation: "We detail the initial spin speed of all of the trials with just the precision necessary to entail in which section the pointer will end. We also give a precise description of the causal mechanisms which determine how quickly the wheel slows down, and we describe where the boundaries of the red and black sections are. Together, this information entails the outcome of each spin of the wheel; and therefore it entails E."

Is this the best low-level deterministic explanation that is available? In Strevens's book, a good explanation is one that is as abstract as possible. Have we gone far enough in our abstraction of the low-level explanation, or can we make it more abstract without invalidating the causal entailment of E?

Further abstraction is indeed possible. We do not need to give a *precise* description of the causal mechanism that make the wheel slow down: all we need to provide is a mapping of ranges of initial spin speeds to ranges of points on the wheel. The facts we need for the causal entailment of a single outcome have the following form:

1. The initial spin speed in this trial lay between v_n and v_{n+1}.

2. When the wheel spins with an initial speed between v_n and v_{n+1}, it will come to rest between ϕ_n and ϕ_{n+1}.

3. All points on the wheel between ϕ_n and ϕ_{n+1} are red (or black).

Maximum simplicity is attained when the two speeds v_n and v_{n+1} map precisely on the beginning and the end of a single red (or black) section.

The full explanation will not be an ordered list of such small explanations, since the order in which the different initial speeds are realised is irrelevant to

the outcome. Instead, the full explanation will say that N_1 of the initial spin speeds lay between v_1 and v_2, that N_2 of the initial spin speeds lay between v_2 and v_3, and so on. We still have a low-level deterministic explanation of E, but it is much more abstract that Strevens's characterisation implied. For further reference, I will call this explanation LLD.

Can we abstract even further, to get an even better explanation? This is a difficult question, because although it is clear that we can abstract further, it is not clear whether the explanation becomes better. Indeed, one of the main problems of Strevens's theory is that it does not make clear when further abstraction no longer leads to increased explanatory power (Gijsbers 2009, [32]).

For instance, LLD implies a precise number of red outcomes; but in order to imply E, we need only show that the number of red outcomes falls within certain limits. Suppose that we actually 503 red outcomes, and E states that we had between 450 and 550 red outcomes. In that case, we can construct more abstract versions of LLD by taking any 953 spins of the wheel, showing how many red outcomes they lead to, and point out that it is impossible for the remaining 47 spins to make a difference. This procedure may give us up to $\binom{1000}{953}$ explanations, all of which imply E and all of which contain less information than LLD.

But are these explanations really better? Does the existence of such explanations show that LLD contains many irrelevancies? It is far from clear that additional information is always perceived as being irrelevant. Consider the following two explanations of why John cannot drink alcohol: "You must be at least 16 to drink alcohol. John is not yet 16." "You must be at least 16 to drink alcohol. John is 14." Is the first explanation better than the second, because it contains less irrelevancies? I submit that both explanations are equally good, and that information that is not strictly necessary for the entailment does not have to be an explanation-degrading irrelevance.

This topic merits more discussion than we can give it here. Fortunately, for our purposes nothing hinges on whether the best low-level deterministic explanation is LLD or some more abstract variant. Let us call the best low-level deterministic explanation LLD^*.

Aren't we begging the question, by assuming that such an LLD^* (a low-level deterministic model without irrelevancies) exists? After all, Strevens writes that:

> The rationale for the low-level model's irrelevancies is its concern with predicting individual outcomes. As the [indeterministic] explanation shows, in order to predict an approximate frequency, it is unnecessary to predict the outcomes. (Strevens 2008 [132],

p. 381.)

The suggestion is that *any* low-level deterministic explanation of E must contain irrelevancies, since we can get the same goods while bypassing the low level entirely.

It is clear that the argument cannot work as written. It is certainly true that we can *predict* the approximate frequency without predicting any of the individual outcomes; but we are trying to *explain* the fact that in this particular set of trials, the expected frequency was actually attained. And even if we grant that this can be explained while bypassing the individual outcomes, we may still add that a deterministic explanation that does not bypass the individual outcomes is the better explanation. After all, we do *not* get the same goods: LLD^* entails E with certainty, and thus contains all of the difference makers for E; the indeterministic explanation, to be called ID, entails E only with high probability, and misses some of the difference makers.

Given that we have followed Strevens's own procedure of abstraction, no irrelevancies can remain in LLD^*. An irrelevance is something that we should abstract away, and by hypothesis, there is nothing further to abstract away in LLD^*. The only worry might be that LLD^* is no longer a low-level deterministic explanation. But this worry is unfounded: one element of the abstraction procedure is that the causal entailment of E remains intact, and we have already seen in our discussion of the statistical deterministic explanation that only a low-level deterministic explanation can causally entail E.

LLD^*, then, contains no real irrelevancies. But perhaps something can be a difference maker, and hence explanatorily relevant, without its inclusion making the explanation better? This indeed seems to be Strevens's position. He explicitly denies that all difference makers are explanatory, and claims that there can be trade-offs between accuracy and generality (Strevens 2008 [132], pp. 146-148), where accuracy is a measure both of the probability and the precision with which the explanandum is entailed. If dropping a difference maker from a deterministic explanation decreases the accuracy only a little, but increases the generality a lot, then including the difference maker degrades the explanation.

The point of contest, then, is whether there *is* a trade-off between accuracy and generality when we want to maximise explanatory power. I submit that there isn't: all difference makers are explanatory, and the more difference makers we know about, the better we understand the explanandum. (There is of course trade-off between explanatory power and pragmatic factors: if a small increase in explanatory power comes at the cost of making

the explanation much more involved and harder for a finite intelligence to grasp, we will generally forgo it. But this is a completely different point.) If Strevens is right, IC is a better explanation of E than LLD^*, since ID is only a little less accurate, but vastly more general; if I am right, LLD^* is the better explanation.

In the next section, I will discuss Strevens's explicit example of a difference maker that has no explanatory power. In section 8.9, I will then give an argument that shows that his theory has undesirable consequences.

8.8 Non-explanatory critical events

Strevens writes:

> [I]n cases where an event e is best explained probabilistically, a critical event e is explanatorily relevant to e only if it makes a positive contribution to the probabilification of e [...]. It is true, I will contend, even when e is deterministically produced; the thesis is therefore in direct opposition to both the counterfactual and the manipulation approaches to causal explanation. (Strevens 2008 [132], p. 448.)

Strevens uses an example to prove that there are cases where the counterfactual and manipulation approaches give the wrong answer and his theory gives the right answer. Suppose that as I leave a party, I suddenly remember that I haven't thanked the host. I turn back inside, and thank him. Because of this slight delay, on my way home I get involved in an accident with a unicyclist. If I hadn't thanked the host, I would have been at the fateful location five minutes earlier, and I would not have hit the unicyclist.

Thanking the host is a difference maker to the collision. But Strevens's suggestion is that it depends on the context whether thanking the host is explanatorily relevant to the collision.

Context I The probability of an accident of this type was low.

Context II The probability of an accident of this type was high. I was drunk, I am a bad driver, the weather was terrible, and there was a huge unicyclist convention going on. (Still, my thanking the host *was* a difference maker – had I not thanked him, I would have had the good luck to avoid all unicyclists.)

According to Strevens, in the first context the best explanation is deterministic, and any critical event (any difference-making event) will appear in the deterministic model. Thus, in the first context, the thanking of the host is explanatory. But in the second context, the best explanation is an indeterministic model, and this model will not mention the thanking, since it will not mention the exact time that I left. Strevens continues:

> [T]his doctrine accords with practice. It would be strange [in the second context] to explain the accident by noting that I stopped to thank my host. [...] In the case where there are no probability-raising factors, by contrast, it is not strange to cite my stopping to thank the host [...]. The explanation goes as follows: it was just a matter of unfortunate timing, the unlucky combination of events such as my stopping to thank the host, my getting temporarily trapped in the guest bathroom, and so on. What makes the citation of these events relevant is the truth of the preamble: it was *just* a matter of unfortunate timing. When this claim is false – when there are significant probability-raisers such as the unicyclists' convention – the little matters of timing are no longer a legitimate part of the explanation. (Strevens 2008 [132], p. 450)

Strevens emphasised the word *just*, and he should. For the accident *is*, after all, a matter of unfortunate timing, even in the second context – it's just not *just* a matter of unfortunate timing. Surely it is legitimate to say in one's explanation that the accident is a matter of unfortunate timing when it is in fact a matter of unfortunate timing, even if it is also a matter of many other things, such as my driving skills and the number of unicyclists in town?

Suppose I am telling you about my evening, and I produce a chart that shows which departure times lead to a collision with a unicyclist and which do not. Assume that most of the chart is coloured the red of 'collision' – we are in context II here. Now I am telling you about my evening, and I point to a white portion of the chart and say: "And that's when I got into my car." "Lucky chap! That was exactly the right time to leave!", you exclaim. "Well," I say, "it would have been. But then I remembered that I had forgotten to thank my host, so I went back inside...". And my finger moves to a big red area five minutes further on.

This exchange *will* give you understanding of why I collided with a unicyclist. I show you that the chance of a collision were very high, but that I had nevertheless *almost* escaped by a feat of incredible timing. Unfortunately, my remembering to thank my host made the difference to the occurrence of

this difference-making feat of good timing; so I tell you about the thanking of my host, and you understand why I had a collision after all.

"I don't understand how you could have had an accident. According to my calculations, you left at exactly the right moment." "Yes, but then I went back inside to thank the host." "Ah, now I understand."

Perhaps what is going wrong in Strevens's account is that he makes too naive a use of the idea of probabilistic relevance. The probability of my colliding with a unicyclist can be calculated, of course, but it depends on the information we have about the potential difference makers. Given the place and velocity of all unicyclists in town, my bad driving skills, my level of intoxication and the visibility, the probability of a collision may be very high. But add to this the fact that I was stepping into my car at the precise moment t_1, and the probability of a collision may suddenly be very low. If *that* is the background knowledge from which we reason, my remembering to thank the host has a huge probabilistic relevance. Such a set of background knowledge against which the event has a huge probabilistic relevance can necessarily be constructed for any difference maker – otherwise, it would not be a difference maker. In this respect, there is absolutely no difference between my thanking of the host and my driving skills.[5]

Thus, even if we wish to equate explanatory relevance to probabilistic relevance, there is a sense in which all difference makers are highly relevant. This fact may be obscured by looking at examples in which the 'wrong' set of background information has been given; but it is a fact nonetheless. All difference makers are relevant.

8.9 Deductive closure

In the previous section, I have tried to show that Strevens's example of a difference maker that is not explanatorily relevant is not convincing. I now want to give a general argument against his claim that ID is superior to LLD. The premise of my argument is that the set of propositions for which a potential explanation X is an actual explanation (or, if you prefer, a good explanation) is deductively closed. So if X is a good explanation of E, and E logically implies E', then X (together with the statement that E implies E') is also a good explanation of E'. I will call this the *principle of deductive closure*indexPrinciple of Deductive Closure. (As we will see, the principle is

---

[5]In private communication, Michael Strevens has pointed out to me that he is thinking of physical rather than epistemic probabilities. This complicates the issue, but may not invalidate the argument I give: we would have to talk about ensembles rather than background knowledge, but the relativity is still there.

incomplete as it stands; a full version is given at the end of the paragraph.)

This premise is very natural. If we have a good understanding of why some event E happened, then, provided that we understand the deduction, we surely also have a good understanding of why its logical consequences happened. If we understand why Smith has a daughter, we understand why Smith has a child. If we understand why *Gravity's Rainbow* is one of the best novels of the twentieth century, we also understand why it is one of the best novels of its decade. If we understand why the ball moves with a constant speed of 1 meter per second, we understand why it has moved twenty meters after twenty seconds. And so on.

Still, we need to exercise some caution, because explananda are not facts but facts within contrast classes (Van Fraassen 1980 [28]; this thesis, chapter 5). For this reason it is somewhat misleading to say that if we understand why the man called 'Obama' won the 2008 US presidential elections against John McCain, we understand why a man won the 2008 US presidential elections. Even though "a man won the 2008 US presidential elections" is a logical consequence of "a man called 'Obama' won the 2008 US presidential elections against John McCain", the former is likely to invoke the idea that we are explain why a man, rather than a woman, won; the latter that we are explaining why Obama, rather than McCain, won. And of course explaining the latter is not equal to explaining the former.

For the principle of deductive closure to kick in, the contrast class of E must contain the contrast class of E'. (Here I use 'contrast class' to denote the set which contains both the fact to be explained and its foils.) What does this mean? The contrast class of E contains the contrast class of E' just in case every instantiation of an element of the contrast class of E' is also an instantiation of an element of the contrast class of E. If E' is "Obama (rather than McCain) won the elections" and E is "Obama (rather than anyone else) won the elections", the contrast class of E contains that of E_2: every instantiation of "McCain won the elections" is also an instantiation of "someone not Obama won the elections", and every instantiation of "Obama won the elections" is obviously also an instantiation of "Obama won the elections".

It is easy to understand why this additional criterion must hold for the principle of deductive closure to kick in: the contrast class furnishes a premise for the explanation. If we are asked why Obama rather than McCain won, we are allowed to take for granted that nobody else won. But if we are asked why a man rather than a woman won, we are allowed to take for granted only the much weaker claim that the elections were not won by a gender-neutral or ungendered being. If we can prove p from q, we can then conclude that we can prove r from s provided that (a) p is at least as strong as r, and (b)

s is at least as weak as q.

With this caveat in mind, let me repeat how natural it is to believe that if we have a good understanding of why some event E_1 happened, we also have a good understanding of why any of its logical consequences happened (provided that we understand the deduction). Given that I understand why Obama won, rather than anyone else (including all individual women), I surely understand why a man won, rather than a woman. This is true even if (say) gender discrimination were one of the causes of Obama's victory: if gender discrimination made it the case that a man rather than a woman won, then it must also be the case for at least one woman (Hillary Clinton, perhaps) that she did not win because of gender discrimination, and therefore gender discrimination will appear as a cause of Obama's victory in the first explanation as well as in the second.

Strevens's account of explanation, however, is incompatible with the principle of deductive closure. This is very easy to see. According to Strevens, we can have good understanding of why trials 1 to 500 on a wheel of fortune all came up red; we can have good understanding of why trials 501 to 1000 all came up black; and yet have only a very bad understanding of why approximately half the trials came up red. This is because Strevens denies that good explanations for the first two facts (which will necessarily be low-level deterministic) can form a good explanation of the third fact (since a combination will also be low-level deterministic, whereas a good explanation of E must be indeterministic). Strevens's claim that LLD is not a good explanation of E thus commits him to a denial of the principle of deductive closure.

Indeed, on Strevens's account, we can have good understanding of why *exactly* half of the trials came up red, while having only very imperfect understanding of why *approximately* half of the trials came up red. But surely, getting the latter understanding is strictly easier than getting the former? Saying that LLD perfectly explains why 503 of the 1000 trials were red, but at the same time denying that LLD is successful at explaining why approximately half of the trials were red, is very strange. (And please note that in these examples, all the conditions on contrast classes formulated above are fulfilled.)

The example we have seen in the last section also shows that Strevens claims that a good explanation of why I collided with a specific unicyclist (rather than not colliding with that unicyclist) need not be a good explanation of why I collided with *a* unicyclist (rather than not colliding with one). Again, this is like claiming that I can understand why there are fewer than a thousand countries in the world without understanding why there are fewer than a million.

Strevens counters (in private communication) that although it is true that

if you can explain E you can also explain any deductive consequence E' of E, it need not be the case that the first explanation is identical to the second one. For the first explanation may contain details that are difference makers for E but not for E', and mentioning them as if they were difference makers might invalidate the explanation. Thus one can hold on to the claim that LLD^* perfectly explains why 503 of the 1000 trials were red, but cannot explain why approximately half of the trials were red, because it contains many irrelevant difference makers.

The point is logically sound, but I am not very worried about irrelevant difference makers: since full explanations give us the relations of determination, it should always be obvious which of the determining bases actually make a difference. If one required explanations to make explicit which of their statements give us difference makers (which I do not), then we could not use LLD to explain why half the trials were red; but we could use LLD', which is LLD plus this explication, to do so. The explication would consist of a statement saying how many of the low-level facts would have to be changed in what way in order for the number of red trials to deviate significantly from 500. Not only would LLD' still be a low-level deterministic explanation, it would arguably be essentially identical to LLD.

But the opposite, namely that the explanation of E *lacks* difference makers for E', seems to me a real problem. Let us look once again at the contrast class. Suppose we have an argument that explains why 503 of the trials were red (rather than this not being the case). Such an argument could take the form of citing one difference maker which made the difference between 503 and 504 red outcomes. This would not be an explanation with great strength, but it would be an explanation. However, it could not be adapted to explain why approximately 500 of the outcomes were red (rather than this not being the case), since it contains no difference makers relevant to this contrast. In order to avoid such problems, we need to add a requirement to the principle of deductive closure, which I will write out again for clarity:

Principle of Deductive Closure : If X is a good explanation of E (rather than any other member of D_E), and E logically implies E', then X together with the statement that E implies E') is also a good explanation of E' (rather than any other member of $D_{E'}$), given that the following two condition hold:

1. the contrast class of E' is contained in the contrast class of E; that is, any instantiation of one of the member of $D_{E'}$ is also an instantiation of one of the member of D_E;

2. at least one combination of elements in the determining bases of

X changes the probability distribution over $D_{E'}$ in such a way that the probability of E' is no longer 1.

Our new second condition is easily satisfied by LLD, because its low-level deterministic description of the trials contains all the difference makers one could ever want. So I would argue that the principle of deductive closure, although it has some limits, holds in the situation where Strevens claims it does not.

I have tried to show, first, that the low-level deterministic explanation does not contain any irrelevancies; second, that Strevens's explicit example of a critical event that is not explanatory does not work; and third, that Strevens's claim that the low-level deterministic explanation is not a good explanation leads to the undesirable consequence that the principle of deductive closure must be abandoned. Together, this gives enough reason to hold on to the thought that all difference makers are explanatory and that the low-level deterministic explanation has the greatest explanatory power.

8.10 Conclusion

Michael Strevens argues that a deterministically produced statistical event can be explained in three ways: by a low-level deterministic explanation, by a statistical deterministic explanation, and by an indeterministic explanation. The low-level explanation can be abstracted to the statistical deterministic explanation, it contains many irrelevancies and must be judged inferior. The indeterministic explanation is the best of all, since it is more general than the statistical deterministic one, while being only a little less accurate.

My counterargument consisted of two stages. In the first stage, I have shown that the statistical deterministic explanation does not exist. It therefore no longer follows that the low-level deterministic explanation is inferior, because it no longer follows that it contains irrelevancies. In the second stage, I have argued that the low-level deterministic explanation is superior to the indeterministic one. We first discussed an example used by Strevens to show that critical events can be non-explanatory. I suggested that no such conclusion followed from the example. Then we looked at the principle of deductive closure and saw that Strevens's account violates this principle.

My final conclusion, then, is that Strevens has no convincing arguments that show the low-level deterministic explanation to be inferior to the indeterministic explanation. As long as such arguments are lacking, I would like to suggest that the fact that the deterministic explanation mentions more dif-

ference makers is still a *prima facie* reason to see it as the more explanatory of the two.