The Causal Role of Aggregate Properties

Michael Strevens

To appear in C. K. Waters and J. Woodward (eds.),
Causation and Explanation in Biology,
Minnesota Studies in the Philosophy of Science
University of Minnesota Press, Minneapolis.

Abstract

An aggregate property, such as pressure, population, or concentration, is one whose realization is achieved by the joint instantiation of many more or less independent properties. Working for the most part with a biological example, this paper asks: can an aggregate property be causally relevant to a phenomenon if only some aspects of its realization play a direct role in causing that phenomenon, while the rest are, as it were, causal bystanders? I show that it is difficult for any known theory of causal relevance to distinguish aggregate properties that are genuinely causal from those that are not (such as combinations of causally relevant properties with arbitrary, unrelated properties). I then develop a strategy for understanding the relevance of partially causal aggregates, diagnosing it as explanatory rather than strictly causal.

Acknowledgments

I am very grateful to my commentator Ben Jantzen, to the other members of the two Minnesota workshops where the idea for this paper was conceived and where it was ultimately presented, to the editors of this volume, and to Marco Nathan and Laura Franklin-Hall, for their valuable comments.

1. Causation by Concentration

Concentrations, Marco Nathan (2014) proposes, can be causes. That is to say that the proportion of some mixture constituted by a particular type of molecule, in spite of its being a high-level, aggregate property of the mixture, is and is routinely treated as the kind of thing that can cause things to happen. The rate at which a chemical reaction takes place, for example, is typically and rightly regarded as causally explained by the concentrations of the various reactants.

In certain biological cases, Nathan continues—and these are the focus of his paper—only a tiny fraction of the molecules in a causally relevant concentration may be causally active at one time. This is true for chemical reactions in general: at a given time, only relatively few molecules will be actively involved in the transformations that constitute the reaction. But in Nathan's examples the situation is particularly striking.

The aim of this paper is to better understand what is going on in these and similar cases where aggregate properties are attributed some kind of causal power in spite of the apparent causal impotence of the greater part of their realization—in spite of their being, in a certain sense, causally "lazy". Are lazy aggregate properties literally causes or are they causally relevant in some weaker sense? What distinguishes lazy aggregates that are relevant to their effects from other aggregate properties that are not? By developing a general principle for attributing explanatory power to aggregates, I hope to answer these and related questions.

Nathan's principal example is the "genetic switch" that regulates the behavior of the λ phage, a virus that attacks *E. coli*. When the phage is switched into its lysogenic state, a quiescent state in which the bacterium continues to reproduce normally but with the phage's DNA now embedded in its own, it is kept on track by a handful of molecules of the protein cI ("clear 1") bound to certain operator sites on the phage's DNA. There are six such sites in all, so six molecules of cI are at any moment controlling the chemical course of things

by initiating and suppressing the various reactions needed to maintain the lysogenic state.[1]

In particular, the presence of a cI molecule at one of the sites activates the production of more cI molecules, thus maintaining a high concentration of cI molecules in the environs of the phage's DNA. As a consequence of this high concentration, if one of the bound cI molecules is knocked off its site in the molecular hurly-burly, as typically happens many times each second, it is very likely to be replaced by another molecule of cI, maintaining the switch in its lysogenic setting. (Were it to be replaced by a molecule of the protein cro, present in much lower concentrations in the lysogenic state, that would be a first step toward a flipping of the switch to the lytic state, in which the infected bacterium manufactures many copies of the phage and then ruptures, releasing them into the wide world.)

Nathan claims that both scientists and ordinary people conceive of the high concentration of cI molecules as causal: we happily say that a particular infected bacterium is driven to travel the lysogenic life cycle in part by the high cI concentration. Yet at any time only six molecules of cI are in the driver's seat. These constitute a minuscule fraction of the molecules that make it the case that the concentration of cI is high; thus, the high-level property to which we attribute the causal power to maintain the lysogenic life cycle appears to be almost entirely, over any short interval, causally inert. It is as though we attribute to the entire population of New York City the speed of the 2.13 PM F train, though only a single motorman has the speed controller in their sweaty grasp. How can an aggregate property be causal, when almost none of what it aggregates is doing any causing?

1. More exactly, at most six molecules of cI are in control, since sites may be vacant. The sites have differing functions; one, for example, prevents the concentration of cI from becoming too high. Should a cI molecule's being bound at a suppressor site like this be considered a cause of lysogenic activity? That is a tricky question; it will not be considered here. In the main text, I will not discriminate among the sites, and will talk rather simplistically as though it is six molecules of cI bound to the six sites that jointly bring about lysogenic activity.

* * *

The case of chemical concentrations is, though significant in itself, a part of an even bigger picture, both in biology and elsewhere, of lazy aggregates doing explanatory work.

First, as Nathan observes, there are "concentrations" that are not chemical or biochemical. In frequency-dependent selection, the concentration of a trait affects the rate with which that trait is passed on to future generations. As in the chemical case, we attribute causal responsibility to the frequency itself, though (depending on the details of the case—mimicry is an example) not every possessor of the trait may actively contribute to the selection process in each generation. In economics it is said that, when conditions are right, greater demand causes an increase in supply, though the factory owners may base their decision to speed up the production lines on a survey of demand at only a small sample of outlets.

Second, there are many other properties in the high-level sciences that are ascribed causal or explanatory power, though only some small part or aspect of the properties' realizations does the causing. One example is the attribution to species membership of characteristic traits such as color or territoriality or reproduction rate. We say that a bird is black because it is a raven. If this "because" is understood causally, the attribution is a puzzle, because the bird's ravenhood consists at least in part (on most views) in its historical relationship with certain other birds, and this relationship is causally inert. So, even if some aspects of being a raven are physiological and able to play a part in causing blackness (or reproduction rate, or whatever), some aspects are not, yet we treat the property as a whole as though it is causal or at least explanatory. Another much-discussed example is that of semantic properties, which seem to depend in part on facts outside the head but which are nevertheless attributed a causal role in thought, a causal process that is usually supposed to take place entirely within the skull. There are many examples in the social sciences: causal consequences are variously attributed

the properties of being married, or Catholic, or Australian, or a youngest child, though only a proper part of these properties' realization can be doing the causing in question.

Third and finally, the concentration case is similar in certain important ways to a kind of explanation found in statistical physics. An ice cube is removed from the freezer and placed in an old-fashioned cocktail. It melts. Why? Because it was put in a liquid at room temperature, of course. But why is being placed in a liquid at such a temperature relevant? It is because almost any set of initial conditions realizing an ice cube sitting in a liquid at room temperature will lead to melting. Now of course, only one set of initial conditions drives the causal process that constitutes melting—the actual initial conditions. But somehow the other, merely possible initial conditions contribute to the story by making melting, given that the ice was in a certain macrostate, overwhelmingly probable, thereby rendering the macrostate itself causally relevant. The explanation of a simple case of trait fixation in evolutionary biology has much the same structure—but perhaps I should stop there.

A full account of the causal role of aggregate properties would treat all these cases and their variants. This paper is not so ambitious. Its focus is concentrations, broadly understood, and above all those telling cases, like that of the λ phage, in which the lion's share of the aggregate sits around licking its paws and doing nothing very helpful—the lazy aggregates. Nevertheless, the outlines of the story I give are not confined to concentrations or even to lazy aggregates; I hope, then, to make some progress on the question of aggregate causation in its most general form.

In sections 2, 3, and 4 I do my best to understand the explanatory role of lazy aggregates in genuinely causal terms. I fail. Thus I turn to the alternative strategy, in section 5, of understanding lazy aggregates as explanatory but not literally causal. Here it is possible to gain some traction. It remains very difficult to draw a principled line between aggregates that are explanatory in

spite of their laziness and those that are, precisely because of their laziness, non-explanatory. But I take my best shot.

A methodological note before I jump in: the problem of lazy aggregates arises most clearly for singular causation (also known as token causation or, more recently, as actual causation). That is how I have framed it above and how, for the sake of philosophical clarity, I will continue to conceive of it. Singular causation matters in the law courts and in everyday thinking, but does it play a significant role in science? Yes, for two reasons.

First, some branches of scientific inquiry, notably in the historical sciences, have among their ultimate goals the elucidation of singular causes. What were the causes of the Permian extinction? The determination of the weak mixing angle? The dominance of Indo-European languages across Europe and parts of Asia? The English Civil War?

Second, successful experiment and observation in science depend on a grasp not only of the causal structure of the relevant measurement apparatus, but of singular causes operating in particular instances of measurement. In 2011, for example, readings suggested that neutrinos generated at CERN traveled to central Italy at a speed faster than that of light. Relativity theory was saved when it was determined that the cause of the measurements in question was a loosely connected cable. There were many measurements, but the role of the cable was as a singular cause of the error in each case. The fruitful deployment of almost any experimental setup requires many such judgments about singular causation every day.

2. The Relevance of the Bystanders

That a lazy aggregate property such as a concentration is causally relevant to a process suggests the relevance not only of the few molecules that are actively engaged, at any time, in turning the chemical cogs that drive the process, but also the causal relevance of the rest—the molecules that are simply standing by watching while the minority do the hard work. How are these molecular

bystanders contributing to the process? How are they *causally* relevant?

Let me begin by quickly dismissing two possible answers to the question of bystander relevance. First, you might observe that the bystander cI molecules are not completely causally inert. They are bouncing around in the environs of the λ phage's DNA, and will occasionally strike a molecule that is bound to one of the operator sites, or that at some future time will become bound to one of the sites. In a certain sense, then, they are causally connected to, thus relevant to, the phage's activity. But this cannot be the kind of causal connection that qualifies the bystanders, or the concentration they realize, as a cause of lysogenic activity. Molecules of the protein cro have roughly the same physical properties as molecules of cI, and so high concentrations of cro will bounce with just as much vigor as high concentrations of cI. If this bouncing is sufficient to make a high concentration of cI a cause of lysogenic activity, then it is sufficient to make a high concentration of cro a cause of lysogenic activity—yet a high concentration of cro is not a cause but an inhibitor of lysogenic activity. Going a little deeper, for a causal influence to count as a cause of a process, that influence must make a difference to whether or not the process occurs. But bouncing is quite neutral with respect to lysogenic activity: it neither promotes nor inhibits it. (If anything, it inhibits any kind of stable activity, because it is apt to knock molecules off operator sites.)

Second, you might argue for the causal relevance of the bystanders on the grounds that they may eventually bind to the operator sites themselves, if the lysogenic cycle goes on for some time. This solves the problem, perhaps, if you wait long enough that almost every cI molecule has had a turn at binding. But we can choose a period of time much shorter than this and ask: is the concentration of cI a cause of lysogenic activity over this shorter time period? The answer seems, intuitively, to be yes, but over such a period the vast majority of cI molecules will have remained mere bystanders.

The beginning of the correct answer to this question of the relevance of the bystanders is surely that, if they are relevant at all, then as Nathan remarks

it is in virtue of something like probabilification: a high concentration of cI is a sustaining cause of the λ phage's lysogenic life cycle because the high number of cI molecules relative to cro molecules makes it very probable that it will be cI molecules and not cro molecules that bind to the relevant sites. The relevance of the bystanders, then, lies in their making it very probable that those sites will be filled by some cI molecules or other, and so that the lysogenic cycle will continue.

Is this probabilistic relevance a kind of causal relevance, however? There is good reason to think not. As Nathan goes on to observe, the rest of the molecules probabilify the effect because they are backup causes, that is, things one of more of which would likely have caused the effect if the actual causes had not. And though the term 'cause' is a part of the term 'backup cause', backup causes of an effect are famously, definitively, notoriously not causes of that effect.

The philosophical paradigm of backup causation is the following unscientific scenario. Evil pixies Sylvie and Bruno are tossing cannonballs at a rather valuable antique Korean celadon bowl. Sylvie throws first, and her cannonball destroys the bowl. Bruno also throws, and since he never misses, the bowl would have been broken regardless. Sylvie's throw, because her cannonball gets to the bowl before Bruno's, is the actual cause of the bowl's breaking; Bruno's throw is a backup cause, meaning that, had Sylvie either not thrown or missed, it would have been an actual cause of the bowl's breaking.

A standard test for a theory of singular causation is the application of the theory to scenarios of this sort. What is demanded is that the theory correctly detect Sylvie's throw, and correctly reject Bruno's throw, as a cause of the breaking. Behind this test stands, of course, a standard view: backup causes are not causes but merely would-be causes, and so they will be passed over by a good theory of causation. That is what I mean when I say that backup causes are famously not causes.[2]

2. An event can be for separate reasons both an actual cause and a backup cause, but it

A backup cause typically probabilifies the effect for which it is the backup. Suppose that Sylvie typically hits her target 75% of the time, while Bruno hits his target 100% of the time. Then Bruno's throw, for the same reason that it counts as a backup cause for the bowl's breaking, increases the probability of the breaking. Backup causes are therefore non-causal probabilifiers.

The relationship between the unbound cI molecules and the lysogenic cycle is much like the relationship between Bruno's throw and the breaking of the bowl. The cI molecules are backup causes for the cycle, and so they probabilify the cycle, but this probabilification is non-causal. It does not qualify them as causally relevant to the cycle, and so it is left unexplained how an aggregate property that is realized by molecules, 99.99% of which are unbound hence causally irrelevant, could nevertheless qualify as a cause.

3. A Counterfactual Criterion for Causal Relevance

Let me try a different approach. It begins with the notion that it is not the relevance of the bystanders that makes a lazy aggregate causally relevant, but rather the relevance of the aggregate that endows its parts, including the bystanders to some degree, with causal relevance. The idea, then, is that the aggregate as a whole satisfies some appropriate criterion for causal potency; we should stop worrying about the bystanders and seek out this criterion.

Nathan does not aim to provide a general theory of causation by aggregates. But he does suggest a useful starting point by offering the following philosophical rationale for classifying the concentration in the λ phage case as causal. The state of the λ phage's genetic switch—the switch that determines whether it continues in the lysogenic cycle or changes to the lytic cycle—counterfactually depends, he notes, on the concentration of cI molecules, in the following sense:

> Had the concentration of cI molecules been considerably lower, the

cannot be an actual cause in virtue of being a backup cause.

lysogenic activity would not have continued.

It is generally agreed, Nathan goes on to say, that counterfactual dependence of this sort is sufficient for causal relevance (p. 202–203).[3]

That general agreement obtains only, however, when the dependence is between discrete events. When it involves the instantiation of aggregate properties, the picture is rather murkier.

Consider for a moment why the counterfactual above holds, according to a standard Stalnaker/Lewis ("closest possible worlds") theory of counterfactual conditionals (Bennett 2003). To evaluate the counterfactual, you find the possible worlds closest to the actual world where the concentration of cI is "considerably lower" and see what happens there.[4] The counterfactual holds just in case lysogenic activity ceases in all such possible worlds (or at least in almost all such worlds; see note 5). On Lewis's definition of closeness, this is roughly equivalent to the following procedure. First, construct scenarios in which everything about the phage's environment is roughly the same, except that the actual mix of cI and cro has been replaced with a mix containing much less cI and much more cro, changing all else as little as possible. Second, allow these scenarios to evolve according the actual laws of nature, that is, determine what the actual laws predict given the initial conditions that constitute a scenario. If in every scenario, they predict the cessation of lysogenic activity—as they do, since cro molecules will tend to replace the cI molecules bound to the operator sites—then the counterfactual is true.[5]

3. It is not clear to me that Nathan endorses the application of the principle—that counterfactual dependence is sufficient for causation—to just any aggregate. He may view it as valid only for aggregates where every component is a potential initiator of the effect in question.

It is tempting, however, to apply it far more widely. In the philosophy of mind, for example, a counterfactual criterion for causal relevance has often been invoked to establish the causal relevance of semantic properties to thought (Lepore and Loewer 1987). The case is different from that of concentrations in that the causally inert elements of semantic properties, because they lie outside the head, are unlike idle cI molecules not even backup causes for the events putatively caused by the aggregate.

4. For expository simplicity, I am making what Lewis calls the "limit assumption".

5. A philosophical nicety: the laws will predict non-lysogenic activity with high probability,

The counterfactual *is* true, so the high cI concentration passes the test for causality. But why does it pass—why is the counterfactual true? I offer the following diagnosis. When you construct the scenarios in which the counterfactual's antecedent is true, that is, in which the concentration of cI is much lower, you are constructing scenarios in which things that we all acknowledge to play a direct causal role in producing lysogenic activity, namely the particular cI molecules that bind to the operator sites, behave differently than in the actual world or perhaps do not even exist. The counterfactual is true, then, because in the relevant closest possible worlds, the facts actively driving the causal process—the facts about which molecules are bound to the sites—change. That these counterfactual changes result in the cessation of lysogenic activity plausibly tells you that the bound molecules are causes of lysogenic activity. But it is unclear why they should be interpreted as telling you that the concentration as a whole, with its retinue of idlers, is a cause.

To put it another way, the only reason the cI concentration passes the counterfactual test for causal relevance is that a certain aspect of the concentration's realizer (the bound cI molecules) passes the test, and *when you change the concentration you also change the realizer*. To regard the test as showing that the concentration itself is causal, then, would seem to make sense only if it makes sense to hold that a property is causally relevant if any aspect of its realization is causally relevant. But that is absurd—that is what gets you the causal relevance of the population of New York City to the speed of the F train, just because the population includes the person whose hand is on the speed controller.

Indeed, a fine-tuned F train example will show that the counterfactual criterion for relevance is, when applied to lazy aggregate properties, hopeless. (Using an example of psychosocial, as opposed to strictly biological, causation

but not for certain. Let us stipulate that high probability is sufficient for the truth of the counterfactual. (Alternatively, we could stipulate that a counterfactual conditional with a probabilistic consequent is sufficient for causation.)

will allow me to build with minimal effort some scenarios that are similar and different in various ways to the case of causation by concentration.)

Suppose that the F train comes into the 14th Street station too fast and overshoots by a couple of cars. The reason: the train driver checked her voicemail at the previous station and found that she had won the lottery. In her exuberance, she misjudged her speed for the next stop. Allow, then, that the driver's sudden exuberance was a cause of the overshoot. Now look at the average level of exuberance on the train. It underwent a sudden increase at the previous stop (thanks to the driver). Was the increase in average exuberance causally relevant to the overshoot? Intuitively, no: what was relevant was not the increase in the train's total exuberance, but rather the increase in the driver's exuberance. The passengers' exuberance, either individual or aggregate, plays no causal role whatsoever in the story.

Now apply the counterfactual criterion. The relevant conditional is this: had the average level of exuberance not increased, the train would not have overshot. True or false? If we assume that no one else on the train was in a position to learn exciting news immediately before the 14th Street stop then the closest worlds in which the average exuberance does not increase are those in which for some incidental reason—bad phone reception?—the driver's exuberance does not increase. In such worlds, there is no overshoot. The conditional is therefore true, and so average exuberance passes the test for causal relevance. The test, then, issues the wrong judgment about this case, and it is wrong because it is made on precisely the grounds that bothered me above: in the closest world where the aggregate property in question is absent, one of its realizers is absent, and that realizer is a genuine cause.

Observe that the counterfactual criterion is prone to making the same misjudgment even concerning aggregates that are not in any sense unified. Take the average level of exuberance of the group consisting of the train driver and the members of my philosophy of science class (which happened to be in session during the overshoot). Assuming that the students, while in

class, are unlikely to experience sudden episodes of exuberance—a hypothesis that I have on many occasions experimentally confirmed—the closest world in which that artificial, gerrymandered group does not undergo a sudden increase in exuberance is one in which the driver fails to learn of her win, and so in which the train does not overshoot. Thus the overshoot counterfactually depends on the exuberance of the group. You can see how to generalize, I hope: the counterfactual criterion will count infinitely many arbitrarily assembled aggregate properties as causally active when they are in fact quite obviously causally inert.

Can the counterfactual criterion be rescued by a restriction on causation imposed by David Lewis? As I remarked above, Lewis (1973) limits his counterfactual account of causation to discrete events, where what constitutes a discrete event is to be determined by his well-known criterion for "naturalness" (Lewis 1983). He thereby appeals, in effect, to a metaphysical division of the world into natural causal nuggets, each sufficiently well integrated that the causal effects of its parts can be comfortably attributed to the whole. (We say that the bowl was broken by the cannonball, not by the front end of the cannonball.) It is not clear to me whether Lewis would count the existence of a concentration at a time as an event, but it will surely be a formidable task to deliver a metaphysics of naturalness that carves up the universe to include all and only those aggregate properties to which scientists are inclined to attribute causal relevance.

One powerful reason for skepticism is that a solution in terms of naturalness is Procrustean: it does not allow that an aggregate property might count as a single unit for determining the causes of one kind of effect but as many independent units for determining the causes of another. Yet we need this flexibility. For example, although we do not want to count the increase in the average level of exuberance on the F train as a cause of its overshoot, we might very well want to count it as a cause of the increased level of noise on the train after a Knicks NBA Championship win.

4. Three Approaches to Lazy Aggregates

I remain convinced, like Nathan, that in biology and elsewhere, concentrations can have a relevance that arbitrary lazy aggregates lack, even when the vast majority of their realizers are causal bystanders. But so far I have been unable to find a way of thinking about bystanders that, on the one hand, makes sense of the causal relevance of aggregate properties for which they constitute the majority of realizers, and on the other hand, acknowledges that they are not themselves in any sense causes.

As I see it, there are three broad approaches to the problem of understanding the causal importance of lazy aggregates, some of which I have sampled already.

The first is what you might call actualism. Bystander or backup events or entities are not genuine causes, according to the actualist, and consequently, nor are aggregate properties that are realized in large part by bystander events or entities. In a λ phage locked in the lysogenic cycle, then, the causes of the cycle are the molecules bound to the operator sites and no others. An actualist approach owes us an explanation of why we are nevertheless inclined to assert that concentrations are causal. Are we simply confused?

The second approach is counterfactualism. On this view, bystanders may count as causally relevant in virtue of their backup role, that is, in virtue of the fact that if things had gone slightly differently, they would have been actual causes. When they are relevant, the aggregates that they realize are also relevant. A counterfactualist story owes us an explanation how backup causes can be at one and the same time causally relevant and yet not actual causes.

The third approach puts aside the question of the causal relevance of the bystanders, and instead provides a direct test for the causal relevance of aggregate properties. I have already discussed and rejected one aggregate-centered principle at length, namely, the counterfactual dependence criterion tentatively suggested by Nathan, according to which an event or state of affairs involving an aggregate property is a cause of a distinct event if, had the first

event not occurred, the second event would not have occurred.

This paper will eventually advocate another such principle. Before moving on to this positive part of the project, however, I would like to take a closer look at the options for actualist and counterfactualist approaches.

Counterfactualism On a counterfactualist approach, backup causes are attributed, in virtue of their status as backups, a certain causal standing or relevance. Jackson and Pettit (1992) suggest, for example, that backup causes can play a role in a causal explanation if the causal pathway by which they would bring about the event in question is sufficiently similar to the actual pathway. (Jackson and Pettit do not quite go so far as to call backups causes, but they come extremely close, saying that they are "non-incidental features of the actual causal process" (p. 14); I interpret this as an attribution of causal relevance.)

Such a thesis deals effectively with the F train cases: since the exuberance levels of the passengers on the F train, or of the students in my class, are not backup causes of the train's overshoot—roughly, because there are no close possible worlds in which they would bring about the overshoot—neither they nor any aggregates that they help to realize are relevant to the overshoot.

What about genuine backups that are nevertheless irrelevant? The canonical example of such causes is the throw that could have, but did not, break the bowl. Recall the case where Sylvie and Bruno each throw cannonballs at the celadon bowl, with Sylvie's ball being the one that breaks it. Bruno's throw satisfies Jackson and Pettit's criterion for causal relevance: it is a crucial part of a causal pathway that might have led to the bowl's breaking and it is very similar to the pathway that actually led to the breaking. But it is not only wrong to call it a cause of the breaking, it is a mistake to impute to it any causal relevance at all.

Further, aggregate events realized in part by Bruno's throwing do not seem causal either. The simplest such event is the conjoint event of both Sylvie's and

Bruno's throwing. This event occurs (since Bruno actually throws), and it is realized by events all of which are relevant according to Jackson and Pettit. But it seems quite wrong to say that Sylvie's and Bruno's throws together caused the bowl to break. The correct diagnosis is rather that Sylvie's throw was the sole cause.[6]

A second variety of counterfactualism is contained in what might be called a loose manipulationist account of causal explanation. On this sort of view, a property or event is causally qualified to play a part in the explanation of a phenomenon if it helps to answer "what if things had been different" questions about the causal process that brings about the phenomenon. (The "strict" manipulationist account advocated by Woodward (2003), Halpern and Pearl (2005) and others will be considered separately below.) Clearly, any backup cause will help to answer such questions. (What if Sylvie had missed? What if these particular molecules had not bound to the operator site? What if the concentration had been slightly higher?) The good news is that loose manipulationism counts concentration as causally explanatory; the bad news is that it counts the event of Bruno's throwing and aggregates built therefrom as equally causally explanatory. Like other counterfactualist accounts, loose manipulationism is too liberal in its attributions of causal potency.

Actualism Loose manipulationism counts anything that would answer a "what if things had been different" question as causally explanatory; strict manipulationism only counts actual causes as explanatory (and adds that they are explanatory because they answer such questions). Strict manipulationists, then, deny that Bruno's throw plays any part in breaking the bowl. But in so doing, they seem to rule out the relevance of bystander molecules of cI, and

6. Another test case for theories of causality that deal in high-level properties is the disjunction: either Sylvie or Bruno threw. Counterfactual criteria tend to count the disjunctive event as a cause of the bowl's breaking, but as with the conjunction, this lumping together of the two throws into a single cause seems objectionable. I will not discuss disjunctive events in this paper, however, focusing on high-level properties that are more literally aggregative.

so to cast doubt on the relevance of a high cI concentration, since it has few of the trappings of an actual cause. This is the actualist's characteristic dilemma.

Have I moved too fast? A manipulationist about explanation, such as Woodward (2003), will also tend to hold a manipulationist theory of actual causation (though the two are logically distinct). And on such a view, won't concentration count as an actual cause of lysogenic activity?

It is not clear. On the manipulationist view, a quantity is an actual cause of a phenomenon roughly if, had all other potential causes been held constant and only the quantity itself varied (in a certain proprietary sense of "held constant" and "varied" that will not play a decisive role in what follows), then the probability of the phenomenon's occurrence would have varied. Whether concentration passes this test will depend critically on what count as "other potential causes". If you include the six cI molecules bound to the operator sites under this heading, then concentration will fail the test: hold the six molecules in place while varying the concentration (by varying the number of bystander molecules), and there will be no change in lysogenic activity.

But suppose that an aggregate property's realizers are not allowed to count as "other potential causes".[7] Then varying the concentration will sometimes result in a change to the molecules bound to the sites, and so will sometimes result in a change to the phage's life cycle. Concentration (or more exactly, perhaps, the fact of high concentration) will count as an actual cause of lysogenic activity. For the same reason, however, the general level of exuberance on the F train will count as a cause of its overshooting the 14th Street station, and Sylvie and Bruno's both throwing their cannonballs will count as a cause of the celadon bowl's breaking.

Depending on the policy for counting realizers as competing causes, then, the manipulationist package will either count as an actualist or a counterfac-

7. What makes the question murky is that manipulationist causation is defined relative to a model, and the other potential causes are anything that appears in the model. But no criteria are given for deciding what to include in a model when evaluating claims of causal relevance that are not model-relative.

tualist approach to the causal potency of aggregate properties. In the one case, concentration will not be accorded causal-explanatory power. In the other, spurious aggregative events such as the exuberance increase and the conjunctive throw will be accorded causal-explanatory power which they do not deserve. This is not some peculiar defect of manipulationism; it is an illustration of a greater dilemma, a problem that everyone has, of discriminating between concentrations and irrelevant aggregates on principled grounds.

5. An Explanatory Role for Concentration

A Recipe for Explanatory Relevance From the margins of the discussion so far, I take two ingredients from which I will cook up an account of the explanatory role of lazy aggregate properties, including the case of "causation by concentration". The first ingredient is the aggregate-centered notion that, in the sorts of problem cases considered in this paper, aggregate properties explain events by probabilifying their actual causes. The second ingredient is the inkling, which has been bubbling under throughout much of the discussion, that such probabilification should be distinguished from causation proper: a probability-raising aggregate is explanatorily relevant to the effect in question, but it is not a cause—nor is it causally relevant or part of the effect's causal history.

Combining these ingredients yields the following view: the relevance of a lazy aggregate property such as concentration to an effect such as lysogenic activity spans two steps. First, the aggregate probabilifies an event P, in this case the binding of the six cI molecules to the relevant sites. Second, P causes the effect, lysogenic activity. The relation of probabilification is not causal; thus, the aggregate property, though explanatory, is not a cause. The same can be said for those aspects of the aggregate's realization that are not part of P, such as the presence of particular bystander cI molecules: they are not causal, but they are explanatorily relevant to the effect because they participate in the probabilification of P.

The two-step structure I have proposed may also be found in what Jackson and Pettit (1990) call program explanation. Jackson and Pettit say that one property programs for another property if the instantiation of the first necessitates the instantiation of the second. On their view, a property that programs for a cause of an event is itself causally and therefore explanatorily relevant to that event. An aggregate property counts as causing an event, then, if it necessitates a direct cause of that event.

My view departs from Jackson and Pettit's in three ways. First, I substitute a weaker relation for necessitation, enabling aggregates, including concentrations, to qualify more easily as relevant.[8] Second, I do not consider properties made relevant in this way to be literally causes. Third, I add a further component to the story to handle what I will call the problem of relevance. The remainder of this section will be devoted largely to developing the first and third points. The extent to which the second point is a substantive disagreement is then discussed in section 6.

In Strevens (2008, §7.3), Strevens (2012), and Strevens (2014) I develop a relation between properties that I call *entanglement*, which plays the same role in my theory of explanation as Jackson and Pettit's programming does in theirs: when a property F is entangled with a property P, F is explanatorily relevant to the events caused by P. More precisely, if the instantiation of F is followed or accompanied by the instantiation of P, and the instantiation of P causes an event G, then the instantiation of F is in virtue of these two relations, entanglement and causation, explanatorily relevant to the occurrence of G.

Entanglement requires the satisfaction of two conditions. The principal condition is one of robust connection; the other is a condition of relevance. Let me elaborate.

8. Jackson and Pettit suggest at one point that programming requires only probabilification, not necessitation, but it is difficult to determine exactly what they have in mind. My sense is that, even with qualifications, their programming relation is considerably stronger than my substitute.

Robust Connection What does it mean for two properties to be robustly connected?

Ravens tend to be black. But for the reasons given in section 1, this is not because ravenhood itself is a direct cause of blackness. The direct cause of blackness is some mechanism B involving the synthesis of melanin in organelles called melanosomes and the migration of the melanosomes to ravens' feather-and skin-producing cells. The reason that ravens tend to be black is that ravenhood has a robust connection to B-hood. That connection has two aspects. First, there is a statistical connection between ravenhood and B-hood: almost all ravens have B. This is not enough to establish a tendency, however, because a tendency has a modal component: it does not imply only the actual blackness of ravens, but also the blackness of ravens in many other possible circumstances. To the statistical connection must be added a subjunctive connection consisting in the truth of a wide range of counterfactual conditionals concerning particular actual and potential ravens, such as the following:

> If this particular raven had hatched a day later, it would still likely have had B.
>
> If these two ravens had mated, their offspring would likely have had B.

Together, the statistical and subjunctive connections compose the robust connection in virtue of which ravens tend to have B and so tend to be black.

It is precisely the same kind of robust connection that supplies the principal component of the entanglement relation. Two properties F and P are robustly connected in the sense necessary for entanglement, then, if a system's instantiation of F tends to be accompanied by, or followed by, its instantiation of P, in the sense that, first, most Fs have P, and second, a wide range of counterfactuals of the following form are true:

> If this particular instantiation of F had occurred under slightly different circumstances, it would still likely have been accompanied or

followed by an instantiation of P.

Had F been instantiated in this particular system (in fact it was not), that instantiation would likely have been accompanied or followed by an instantiation of P.

Three remarks on the relation of robust connection so defined. First, it is asymmetric; entanglement will inherit this asymmetry. Second, you might think that, if robust connection is to be an analog of probabilification, it ought to take into account base rates, hence that P ought to accompany F more often than it appears in general. Something similar to this requirement will be achieved by the relevance condition to be introduced shortly. Third, the interpretation of "likely" in the conditionals is not important in the context of the present paper. (It is subjected to closer examination in Strevens (2011).)

Returning to the case of the λ phage, observe that there is a robust connection of the sort just defined between a high concentration of cI near a phage's DNA and the occupation of the operator sites by cI molecules. First, most λ phages surrounded by a high concentration of cI have their operator sites filled by cI molecules. Second, it is true of most λ phages surrounded by a high concentration of cI that, even if things had gone slightly differently (without altering the concentration), the operator sites would still likely have been filled by cI. Third, it is true of most λ phages surrounded by a low concentration of cI that, if the concentration of cI had been high, the operator sites would likely have been filled by cI. (The "most" riders are permitted because a robust connection between properties requires only that a "wide range" of counterfactuals of the given form hold.)

The conditions for the existence of a robust connection are much weaker than the conditions for across-the-board probabilification (let alone necessitation), because they require probabilification only in the nearby possible worlds that are relevant to determining the truth of the pertinent counterfactual conditionals. The raven case illustrates this point nicely. The across-the-board (that is, unconditional) probability of a raven's having the blackness-producing

mechanism *B* might be very low if it was rather unlikely that ravens would evolve a black coloration rather than, say, a brown or pied coloration. This low probability does not undercut the robust connection between ravenhood and *B*, however, because such a connection requires an association between the two properties only in close possible worlds, all of which are worlds where ravens have roughly the same evolutionary history as in the actual world.

What determines which worlds are close? The relevant conditionals themselves. The truth conditions for counterfactual conditionals require you to look to possible worlds with histories as close as possible to the history of the actual world up until a time shortly before the counterfactual antecedent occurs. In the case of ravens, the conditionals have the form: *if such and such had happened to such and such ravens then* ____. Close worlds will therefore share the history of the actual world until just before such and such happened—until just before the time that a raven hatches a day later, or two ravens counterfactually mate, for example. The evolution of the raven species predates all such antecedents, thus it is held constant when evaluating the conditionals.

More generally, when determining whether or not there is a robust connection between two properties *F* and *P*, the relevant conditionals concern the kinds of systems—ravens, λ phages or their hosts, and so on—in which *F* is instantiated in the actual world. The antecedents of such conditionals occur after the systems have first come into existence, and so when determining the truth of the conditionals the actual history of the genesis of the systems—in the biological case, their evolution—will be held fixed. When applied to biological cases, then, the notion of robust connection does not require some special, biological clause explicitly requiring respect for evolutionary history. Such respect simply falls out of an entirely general account of closeness. (Contrast the spirit of this approach to Weber's paper in this volume advocating a special notion of "biological normality", to be used to limit what counts as a relevant possible world when evaluating counterfactuals in biology, or Beatty

et al.'s special notion of historical possibility, used to distinguish "historically realistic" from unrealistic counterfactual scenarios.) .

To sum up: a necessary condition for the entanglement of F and P, and so for the explanatory relevance of an instantiation of F to the effects of an instantiation of P, is a counterfactually robust connection between F and P. This is a much weaker, and so much more liberal, condition than Jackson and Pettit's "programming".

Relevance That a counterfactually robust connection is not sufficient for explanatory relevance, hence not sufficient for entanglement, is demonstrated by an old and familiar example from the philosophy of explanation. Say that a thing is hexed if you wave your hands over it and intone some "magic" words (Kyburg 1965). Now suppose that a certain specimen of λ-infected *E. coli* in the lysogenic cycle is hexed. The lysogenic activity continues. We do not want to say that the hexing is a part of the explanation of the continued lysogenic activity. But there is a robust connection between hexed high cI concentrations and the filling of the operator sites by cI molecules. So without any further restriction on entanglement, we would be forced to say that the lysogenic activity can be explained by the hexed, high cI concentration. (The same objection can be made to Jackson and Pettit's program explanation: if the instantiation of F necessitates the instantiation of P, then a fortiori the instantiation of hexed F-hood necessitates the instantiation of P.)

Intuitively, the hexing is irrelevant to the continuation of lysogenic activity because it is irrelevant to the filling of the operator sites by cI molecules. Or more generally, hexed F-hood is irrelevant to P-hood (for just about any P) because the hexing plays no role in bringing on the instantiation of P. What must be added to the definition of the entanglement of F and P to solve the relevance problem, then, is a requirement that F include no properties that are irrelevant in this sense—that is, no properties that fail to contribute to the instantiation of P. But how, exactly, to define relevance?

Is relevance causal relevance? To require, for the entanglement of F and P, that all aspects of F be causally relevant to the instantiation of P would certainly exclude hexing, but it would exclude lazy aggregates such as concentration as well, for reasons that are by now only too familiar: most aspects of the realization of concentration are causal bystanders.

What about probabilistic relevance? A high concentration of cI increases the probability that cI molecules fill the operator sites, but hexing has no impact on this probability at all. Good, but not quite right for two reasons. First, we do not care about probabilification in general but only in the close possible worlds implicated by the counterfactual conditions that characterize robustness. Second, an account in terms of probabilification is insufficiently general, since in many cases, the robust connection between F and P is not strictly speaking probabilistic, as in the case of the connection between ravenhood and the blackness-producing mechanism.

These considerations suggest a counterfactual criterion for relevance. The most straightforward such criterion is as follows: Suppose that there is a robust connection between F and P, and that H is one component of F. (So: F might be hexed ravenhood, H might be hexedness, and P might be the raven coloration mechanism.) Then H is irrelevant to the robust connection between F and P just in case, in almost all the actual and counterfactual scenarios in virtue of which F and P are connected, P would have been present even if H had been absent.

Consider for example, the sort of conditional in virtue of which hexed ravenhood is robustly connected to the blackness-producing mechanism B:

> If this particular hexed raven had hatched a day later, it would still have had B.

Suppose that the raven had hatched a day later but had not been hexed. It would still have had B. So hexing is declared irrelevant, and is removed from F. Ravenhood alone bears the explanatory relation of entanglement to B-hood.

This example makes the criterion look good, but it is too strong: it will tend to strip lazy aggregate properties of all of their non-causal parts. Consider the instantiation of a high concentration of cI, for example. Take some bystander molecule of cI, that is, a molecule that helps to realize the high concentration but that is not itself at any point attached to one of the relevant operator sites. The presence of this molecule will fail the test for relevance: under most conditions, had it not been present, the sites would still have been filled by cI molecules. So the molecule's contribution to the instantiation of the aggregate must be removed in the same way that hexing's contribution to the instantiation of hexed ravenhood must be removed. When you are done, all you will have left are the six bound cI molecules.

Let me give you at last what I take to be the correct test for relevance. For *F* to be entangled with *P*, I suggest, not only must *F* be robustly connected to *P*, that is, not only must there be a broad pattern of association between *F* and *P*. It must also be the case that the pattern of association is not subsumed by any broader pattern: it must be the broadest pattern available.

Consider, for example, the pattern of association between hexed ravenhood and the blackness-producing mechanism *B*. The pattern is broad enough, but it is subsumed under a still broader pattern of association between ravenhood and *B*-hood: every case, actual and counterfactual, in which hexed ravenhood is accompanied by *B*-hood, is also a case in which ravenhood is accompanied by *B*-hood, but not vice versa. Thus ravenhood, but not hexed ravenhood, is entangled with *B*-hood. The same treatment eliminates hexing from the explanatory story wherever it threatens to intrude.

I have presented this pattern subsumption notion of relevance very informally; many details remain to be supplied. My final few pages will be better used, however, to show how the definition of entanglement incorporating the pattern subsumption account of relevance deals with various problem cases of putative aggregate causation. That will, I hope, give you the motivation to look at the fuller presentation of the pattern subsumption account in Strevens

(2008, §7.3).[9]

Entanglement at Work So: to the 14th Street station. The direct cause of the F train's overshoot is the driver's exuberance. What prevents some aggregate that takes into account that exuberance among other unrelated things from achieving explanatory relevance as a consequence? For the most part, such aggregates will not be robustly connected to the direct cause. An increase in, for example, the average exuberance of everyone on the F train is typically not accompanied, in either the actual or most counterfactual worlds, by an increase in the driver's exuberance (or not one pronounced enough to cause an overshoot).

What if the increase in average exuberance is enormous? Suppose the Knicks do win the championship, and as a result everyone on the train gets very excited, including the driver, who overshoots the station as a result? Such a high level of average increase, you might think, is somewhat robustly connected to an increase in the exuberance of any individual included in the aggregate. So arguably, the general level of jubilation should count as explanatorily relevant to the overshoot, though not as a cause of the overshoot. Counterexample? No; I think that in this case, the average level of exuberance on the train does help to explain the overshoot, in just the way that a high concentration of cI molecules explains lysogenic activity. Indeed, I think that the pattern subsumption criterion requires that you cite not the level of exuberance on the F train alone, but in New York City as a whole. Aggregate properties of the city's population can, then, under certain circumstances explain events on the F train that are caused by the driver alone. It is simply a matter of explanation by concentration.[10]

9. In Strevens (2008) and Strevens (2012) I suggested an additional counterfactual criterion for relevance (superior to the one described above). I now suspect that this criterion is merely a heuristic used to test for satisfaction of the pattern subsumption criterion. Strevens (2014) succinctly summarizes my current view, including the enhancement made in section 5 below.

10. What if the driver did not hear the news of the win, but overshot anyway for some

Back to the case where the driver overshoots after learning of her lottery win. Just before the overshoot, I announce to my philosophy of science class that everyone is getting an A just for turning up to the exam. The level of exuberance in the class soars. Thus it soars also in the group comprising the students and the F train driver. Does this aggregate increase explain the overshoot? No, because an increase in exuberance in this gerrymandered group, no matter how extreme, is not usually accompanied by an increase in the driver's exuberance. (The events that cause exuberance in the group as a whole—the promise of As for everyone, cancelled classes, the refutation of Popperianism—are events that typically have no effect on the driver.) There is therefore no robust connection between the aggregate and the actual cause of the overshoot.

Suppose, next, that the driver is hexed just as she hears the news about the lottery. Hers is a hexed increase in exuberance, then—an aggregate of hexedness and exuberance. Does the driver's hexed exuberance explain the overshoot? No, because the robust connection between hexed exuberance and exuberance (guaranteed as a matter of logic alone) is subsumed by an even broader pattern, the robust connection between exuberance and exuberance (also logically guaranteed). That is a bit tricky of course, but it captures rather nicely the reason that hexing is irrelevant to the robust connection between hexed exuberance and exuberance: you get exuberance from hexed exuberance only because you get exuberance from exuberance. (The same treatment can be applied to other cases where an aggregate property is constructed by conjoining the actual cause with some other event, whether or not it is a backup cause. A pertinent example is the conjunction of Sylvie's throw and Bruno's throw considered as a putative cause of the broken bowl, used to cast doubt on counterfactualism in section 4. There is a subtlety in analyzing this case that would, however, take us a little too far out of the way to spell out.)

---

other reason? Exuberance is not relevant in these cases for a reason that will be given in section 5.

A Disjunction Problem Let me briefly deal with a shortcoming of the pattern subsumption approach to relevance as it has been developed so far. In most cases where cI molecules fill the λ phage's operator sites, the concentration of cI in the environs will be high. But there are other possibilities. Some time in the next century or so, for example, we might be able to build nanobots that fill the sites regardless of the overall concentration of cI, by bringing in individual cI molecules and fixing them in place by hand. In that case, you might think, a high cI concentration is not entangled with the filling of the sites, since the robust pattern of association between the property of there being a high cI concentration in the vicinity and the site-filling is subsumed under an even broader pattern of association, that between the property of there being a high cI concentration or an appropriately programmed team of nanobots in the vicinity, on the one hand, and the site-filling, on the other. It seems wrong, however, to count the disjunctive property, and not the disjunct that is actually realized—either a high cI concentration or a team of bots—as what is explanatorily relevant.

A solution to the problem is proposed in Strevens (2008) §7.34: it is an addition to the pattern subsumption account of relevance, which takes the form of a cohesion constraint that limits the extent to which a pattern of association can legitimately be generalized for the purposes of determining relevance. The constraint requires that the instances of the subsuming pattern exist in almost every case "for the same reason". In the scenario at hand, the reason for the robustness of the connection between the high concentration and the site-filling is sufficiently different from the reason for the robustness of the connection between the presence of the bots and the site-filling that together the two connections do not form a single pattern of robust association.

How to individuate reasons? Typically, causally: the mechanism by which the sites are filled in the presence of a high cI concentration is different from the mechanism by which the bots fill the sites. Note that the mechanisms in question do not necessarily establish a causal connection between the

putatively entangled properties: there can be a causal explanation of the robust connection between F and P without F's itself causing P. It is on this fact that my solution to the disjunction problem turns.[11]

6. Causation by Concentration?

Is there such a thing as causation by concentration, in lazy cases where only a small number of the molecules that realize the concentration are causally active? If what I have said about the causal-explanatory role of concentrations is correct, then the answer is no. A high cI concentration is explanatorily relevant to the causation of a λ phage's lysogenic life cycle, but it is not itself causally relevant—it is not itself a cause.

Why, then, does it seem acceptable to talk causally about the role of the concentration? Why is it not a flagrant error to say that a high concentration of cI caused the phage to continue on its lysogenic life cycle, rather than switching to the lytic cycle?

My answer is not particularly clever: I suggest that this is loose talk. Although in our minds we distinguish genuine causal relevance and the kind of non-causal explanatory relevance borne by lazy aggregate properties such as cI concentration, in our conversation we do not always make that distinction clear—though a hedge often marks the elision, as when Nathan writes "strictly speaking, there is nothing that is 'caused' by concentrations" (p. 199).

Is there a difference between saying, as Nathan, Jackson and Pettit, and others do, that there are two kinds of causal relevance, "direct" or "strict" and "indirect", with concentrations causing only indirectly, and saying as I do that the one is genuine causal relevance and the other is a species of non-causal explanatory relevance that is frequently invoked using causal idioms? On the first sort of view, it is claimed that there is some genuine similarity between the

11. More generally, reasons may be individuated on explanatory grounds. Needless to say, in order to avoid circularity, such grounds must be founded ultimately on explanatory relations other than entanglement.

two relevance relations. On the second view, there is no such similarity; using the causal idiom is metaphorical. More particularly, it is a kind of synecdoche: we call the relevance of a high cI concentration to lysogenic activity causal because the latter part of the concentration's two-step relation to the activity is causal. And that is the only causation at work.

References

Beatty, J., E. Desjardins, and D. Crawford. (Forthcoming). What does it mean to say that "history matters", causally speaking? In Waters and Woodward (forthcoming).

Bennett, J. (2003). *A Philosophical Guide to Conditionals*. Oxford University Press, Oxford.

Halpern, J. Y. and J. Pearl. (2005). Causes and explanations: A structural-model approach. Part 1: Causes. *British Journal for the Philosophy of Science* 56:843–887.

Jackson, F. and P. Pettit. (1990). Program explanation: A general perspective. *Analysis* 50:107–117.

———. (1992). In defense of explanatory ecumenism. *Economics and Philosophy* 8:1–21.

Kyburg, H. E. (1965). Comment. *Philosophy of Science* 32:147–151.

Lepore, E. and B. Loewer. (1987). Mind matters. *Journal of Philosophy* 84:630–642.

Lewis, D. (1973). Causation. *Journal of Philosophy* 70:556–567.

———. (1983). New work for a theory of universals. *Australasian Journal of Philosophy* 61:343–377.

Nathan, M. (2014). Causation by concentration. *British Journal for the Philosophy of Science* 65:191–212.

Strevens, M. (2008). *Depth: An Account of Scientific Explanation*. Harvard University Press, Cambridge, MA.

———. (2011). Probability out of determinism. In C. Beisbart and S. Hartmann (eds.), *Probabilities In Physics*, pp. 339–364. Oxford University Press, Oxford.

———. (2012). The explanatory role of irreducible properties. *Noûs* 46:754–780.

———. (2014). High-level exceptions explained. *Erkenntnis* 79:1819–1832.

Waters, C. K. and J. Woodward (eds.). (Forthcoming). *Causation and Explanation in Biology*, volume 20 of *Minnesota Studies in the Philosophy of Science*. University of Minnesota Press, Minneapolis.

Weber, M. (Forthcoming). Causal selection versus causal parity in biology: Relevant counterfactuals and biologically normal interventions. In Waters and Woodward (forthcoming).

Woodward, J. (2003). *Making Things Happen: A Theory of Causal Explanation*. Oxford University Press, Oxford.