The Large-Scale Joints of the World

Ned Hall †
ehall@fas.harvard.edu

1. Introduction

What is the compositional structure of reality?

That question divides naturally into these two: What is the compositional structure of the *particulars* that populate reality? And what is the structure of the *properties* and *relations* that fix what these entities are *like*?

David Lewis's work in ontology and mereology provides the materials for an extraordinarily clean answer to the first question. First, among the particulars[1] that populate reality are *mereological simples*: entities that have no proper parts. (A plausible candidate for these simples: spacetime points.) Second, every collection of such entities has a *unique mereological fusion*. And third, every particular is either a simple, or a fusion of simples.[2] That's it.

I propose to take this answer on board.[3] What, then, about our second question? Here it looks as though we can draw on an additional Lewisian thesis:

> **Joints**: There is a distinction – at the level of metaphysics – between *more and less natural properties*. Some properties (*having mass 1 gram*, perhaps) are *perfectly natural*; others (*being a methane molecule*, perhaps) are less-but-still-quite natural; still others (*being grue* is a favorite) are not very natural at all.

† Harvard University, MA, USA.

[1] Note that the restriction to "particulars" is in place because Lewis allows that there might be *other* sorts of entities – e.g., repeatable universals.

[2] I said that Lewis's work provides the materials for this answer. It is much less clear whether Lewis himself endorsed this answer; the textual evidence is somewhat ambiguous. At any rate, if he did, he endorsed it as a contingent thesis; see for example the opening section of his 1994. I myself prefer the view that this answer, if correct, is metaphysically necessary but a posteriori; see my 2011a.

[3] Though only for the purposes of this essay: there are, after all, reasonable grounds for reservations. What about *holes*, for example? The gyrations Argle goes through (Lewis & Lewis, 1970) to accommodate them suggest to me that a more relaxed view in ontology is called for, one that agrees that the existence of any non-fundamental particular must be appropriately *grounded* in facts about fundamental entities – viz., mereological simples – without agreeing that every non-fundamental particular must be *composed* of fundamental ones.

> This distinction earns its philosophical keep because of the number and centrality of the philosophical projects that must presuppose it. And to say that this distinction resides at the level of metaphysics is, at least in part, to say that it is not grounded in facts about human psychology.

What Joints tells us, in effect, is that there are *objective joints in nature* that appear at different mereological scales, constituted by the pattern of instantiation of natural properties and relations by the particulars that exist at those scales. That is an attractive picture, but please note that it is nothing more than a picture. Whereas we were able to get a complete and exact answer to the first of our two questions about the compositional structure of reality – thanks to the fact that the version of mereology that our answer drew upon was itself clear and exact – what we have here is no more than a framework for such an answer. It needs to be filled out in at least two ways: we need an account of what naturalness of properties and relations is; and we need an account of how naturalness at one scale fits together with naturalness at other scales.

Now, these accounts ought, I think, to be constrained by the need to make sense of one of the central aims of *empirical inquiry*, especially *mature* empirical inquiry – i.e., *scientific* inquiry. And that aim is to provide us inquirers with *explanations* for why our world behaves the way it does. As it happens, Lewis (1986b) also defended an interesting and important thesis about what such explanations consist in:

> **Causal Explanation**: To explain an event is to give some information about its history of causes. Since causation is both *transitive* – if event A is a cause of event B, and B of C, then A is thereby a cause of C – and *egalitarian* – even background 'enabling conditions' of an event count as causes of it, notwithstanding the oddity of saying so in ordinary conversation – the causal history of a typical event will almost certainly be vast beyond any possibility of full and accurate conveying. And so, in any particular context in which we seek understanding of why some event occurred, pragmatic factors will play a large role in fixing how much and which parts of that event's causal history ought to be highlighted. But that is a quite unexceptional intrusion of pragmatics, and one that ought to make no difference to the philosophical project of saying what sort of information explanatory information *is*. That project is completed – in the case of events, anyway – once we identify information explanatory of them with information about their causal histories.

What I would like to explore in this essay are the prospects for fleshing out Joints and Causal Explanation in a way that makes for a unified package.

Troubles will quickly appear: as to Joints, it will emerge that there is no obvious way to say how the distinctions in naturalness that reside at the most fundamental mereological scales ground such distinctions at higher scales, while preserving the view that these latter distinctions are (like the former) perfectly *objective*. As to Causal Explanation, it will emerge that Lewis has overlooked something of great importance to us as inquirers who seek to understand our world, which is knowledge not merely of the *causes* of some given phenomenon, but knowledge of that in virtue of which the causes *are* causes. Only when augmented by this latter sort of information does a causal explanation of some event achieve the right sort of explanatory *depth* (to borrow Michael Strevens's apt expression; see (Strevens, 2009).)

The troubles for Joints and the troubles for Causal Explanation are, I think, connected. To say exactly why will require a bit of spelling out, so here I will just offer some teasers.

To achieve a decent philosophical account of what explanatory *depth* consists in, we will need to supplement Lewis's idea that explanatory information is causal information with the distinctively unificationist idea that we improve our understanding of our world by finding ways to *organize our information about it in a cognitively effective fashion*. Adding this dose of unificationism yields a view according to which, at least at scales above the most mereologically fundamental, the distinctions that we as empirical inquirers find most explanatorily valuable to draw *derive* their explanatory value in part from the way in which they collectively organize, in a cognitively effective manner, our view of the subject matter that is the target of our inquiry. So it turns out that it is partly a matter of human psychology what makes for an explanatorily valuable distinction (at least, on the assumption that it is in part a matter of human psychology what makes for effective cognitive organization). So, since talk of nature's "joints" *just is* talk of those distinctions in nature grasp of which is essential for explanation, it follows that the joints in nature (at least, at scales above the most mereologically fundamental) are what they are in part because of facts about human psychology.

But lest you think I'm succumbing to the "postmodern forces of darkness" (to use Sider's delightful phrase; see his 2011), let me highlight two important qualifications: first, nothing in what I will argue will suggest that the most *fundamental* joints in nature – the joints that it is the job of fundamental physics to discern – are to any extent of human origin. Second, the conception of nature's joints that I will sketch is perfectly consistent with the view that the

world possesses a rich and completely objective causal structure. It is just that in organizing our view of this causal structure, we must impose on it certain taxonomies. And the way in which we do so – more exactly, the fact that certain ways of imposing taxonomies are *explanatorily better* than others – cannot itself be explained merely by reference to that causal structure: that structure does not, as it were, force upon us, merely by virtue of its internal nature, certain ways of organizing it.

Let me now try to unpack all this. We'll begin by digging a little more deeply into Joints.

2. Natural, non-natural, more and less natural

Let's start by getting a little clearer on what the natural/non-natural distinction *is*, and then reviewing Lewis's case for taking the distinction on board. After that, we can consider the complications that arise from trying to give a philosophical account of how this distinction can come in *degrees*.

The distinction is meant to divide properties and relations into those that are somehow *genuine* – reflecting or constitutive of real distinctions in the world – and those that are somehow *artificial* or *gerrymandered* – reflecting, perhaps, nothing more than an arbitrary, purely conventional decision to use a certain label in a certain way. Here is another way to put the idea. Consider all the entities that there are, or that there possibly could have been; indulge, for the moment, Lewis's odd view that the latter sorts of entities exist in exactly the same sense as the former. (That indulgence will make things simpler; it's not essential.) Consider all the *sets* that can be formed from these elements – the power set of reality, if you like. Some of these sets will group together entities that, somehow, *belong* together: all the actual and possible electrons, say, or all the methane molecules, or all the wombats. Others will fall short of this ideal. There is the set that contains all the electrons, *and* all the wombats. Or the set that contains all the electrons that exist in worlds with at least one wombat, together with all the protons that exist in worlds with no wombats. And so on. Once you see what's going on, you'll see that the *vast majority* of these sets will fail to group together entities that are alike in some genuine respect (and fail much more dramatically than the two foregoing examples – both of which, after all, could at least be described in English). For short: among all the groupings

that, set-theoretically speaking, there are, only a small minority correspond to *real* distinctions in the world.[4]

Why believe in such a distinction among distinctions? Well, it's likely that you already *do* – at least, if you understood the brief exposition given in the last paragraph. But for all that, you might reasonably doubt that the natural/non-natural distinction is *objective*, in the sense that *what it is* for a property or relation to count as "natural" has nothing to do with human psychological responses to the world we inhabit. To begin to rebut this worry, as well as to flesh out our explication of the distinction, we should appreciate two reasons to endorse it that draw on aspects of our ordinary thought and talk about the world that are so intimate and familiar as to readily escape notice.

The first reason has to do with *change*. When Billy falls in love with Suzy, that is a genuine change in Billy; but it is not a genuine change in Suzy, notwithstanding the fact that she goes from lacking the property of being loved by Billy to having this property. So – and this is, of course, a perfectly familiar point in the philosophical literature – not every gain or loss of a property by a thing counts as a genuine or objective change in that thing. But if we maintain (as we should) the idea that every change in a thing *is* a gain or loss of a property by that thing, and that at least some change is a perfectly objective feature of the world, then we need an objective distinction among properties to say which gains or losses of properties count.

Second, while some similarities and differences among entities are no doubt in the eye of the beholder, some are not. Two methane molecules are more similar to each other than either is to a tomato, period. A comprehensive scheme for taxonomizing the items that populate our world that failed to recognize this fact would, whatever its other virtues, fall short in one epistemically crucial respect: it would fail to correctly limn one aspect of the world's structure. If we take this sort of structure to be an objective feature of the world – again, as it seems we should – and we take it to be constituted by the facts about which properties entities share or fail to share, then we need an objective distinction among properties to say which are those whose pattern of

[4] The last few sentences have tacitly restricted our attention to *properties*. But the natural/non-natural distinction applies to *relations*, as well. To handle, say, two-place relations, we should start by considering all possible *pairings* of (actual and possible) entities, and then consider all the sets of those pairings, the vast majority of which will correspond, so the thought goes, to no genuine way in which two things can be related. And so on, for 3- and more place relations.

instantiation fixes the structure of the world. (Compare (Sider, 2011).) (Arguably, the first point, about change, is just an instance of the second: from a suitably exalted space-time perspective, to talk about change is just to highlight temporal aspects of the world's overall spatiotemporal structure.)

If we begin with this intuitive distinction between natural and non-natural properties and relations, we can get part way to Lewis's conception of *perfectly* natural properties and relations simply by maxing out one criterion: whereas sharing of natural properties makes for similarity, sharing of *perfectly* natural properties makes for *perfect* similarity, or *duplication*.

Consider the property of being a methane molecule. Any two things that share this property – i.e., any two methane molecules – will ipso facto be quite similar to one another. But they need not be perfect qualitative duplicates: their internal configurations might differ slightly, or they might be slightly different in composition (say, one contains a carbon-12 atom, where the other contains a carbon-13 atom). By contrast, sharing of perfectly natural properties is supposed to yield perfect qualitative similarity. Thus, Lewis (1983b, p. 27): «Two things are qualitative duplicates if they have exactly the same perfectly natural properties».[5]

But the way in which sharing of perfectly natural properties grounds facts about similarity goes beyond the requirement that two objects that instantiate exactly the same ones are perfect qualitative duplicates. Consider two Newtonian point-particles that are perfect duplicates, having, say, exactly the same values for mass and charge. Suppose that, in some appropriate units, each particle has mass 1 and charge 1. Now, *having mass 1* and *having charge 1*, let us agree, are examples of perfectly natural properties. By contrast, here are some of the properties the particles instantiate that fall short of being perfectly natural (on the grounds that the sharing of them does not count as a way of being genuinely similar): *having mass 1 or charge 2*, *having mass 2 or charge 1*, *having mass 1 or charge 1*. But, on the assumption that it is metaphysically impossible for a particle to have two distinct values of mass or charge, the

[5] It's clear in the context that Lewis intends the "if" to be understood as "iff". Also, a more careful formulation of the idea would be the following: Two things A and B are qualitative duplicates iff there is a one-one mapping between those parts of A (including A itself) that instantiate perfectly natural properties or relations and those parts of B that do so, such that whenever some part or parts of A (respectively, B) instantiate some perfectly natural property or relation, the corresponding parts of B (respectively, A) instantiate the very same property or relation.

sharing of these three properties guarantees perfect duplication (at least, with respect to mass and charge). In short: if we take the *complete intrinsic nature* of a thing to be what is shared between it and its perfectly duplicates, then the notion of "perfect naturalness" imposes an additional structure on these complete intrinsic natures. Thus, the intrinsic nature of one of our particles is constituted by its being a point particle, having mass 1, and having charge 1; not (e.g.) by its being a point particle, having mass 1 or charge 2, having mass 2 or charge 1, and having mass 1 or charge 1.

Furthermore, the perfectly natural properties are supposed to collectively constitute a kind of minimal supervenience basis for all of reality: that is, the *whole truth* about the qualitative structure of the world is supposed to be grounded in the pattern of instantiation of perfectly natural properties and relations, in a way that involves no redundancy. This requirement provides an extra reason for excluding such "disjunctive" properties as *having mass 1 or charge 2* from the ranks of the perfectly natural, as well as a reason for so excluding "conjunctive" properties such as *having mass 1 and charge 2* (which could have gained membership, if our sole criterion concerned whether sharing the property is a way for two things to achieve a sufficiently high degree of similarity).

A few observations about Lewis's distinction will be helpful for what follows. First, the philosophical importance of the distinction is not limited to its uses in analytic metaphysics (or the theory of reference – more on this, in a moment). It appears, in addition, to be crucial for articulating what is arguably one of the central aims of *physics*, which is to provide an inventory of the fundamental physical magnitudes of our world. For while Lewis takes it to be a job for philosophy to defend the claim that some properties and relations are perfectly natural, he rightly takes it to be a job for physics to figure out which perfectly natural properties and relations happen to characterize our world. Second, notice that as soon as we grant that physics *has* this job, we can see that the distinction we want is not, or at any rate should not be limited to, a distinction among properties and relations: at least for the purposes of physics, it should be seen as a distinction among *determinable magnitudes*. (For more on this point, see (Hall, 2010).) Third, it is an extremely plausible thesis – but not, I think, an indisputable one, given that the thesis is also, ultimately, *empirical* – that perfect naturalness of properties, relations, and magnitudes is closely connected to mereology, via the thesis that the only entities that genuinely instantiate perfectly natural properties, relations, or magnitudes are

mereological simples (perhaps, just points of space and time). There may be exceptions – for example, it may be that the topological structure of space-time is best understood as a perfectly natural feature of *it*, not reducible to perfectly natural properties and relations instantiated by its ultimate parts. But for purposes of this essay I will simply bracket this issue, and take for granted that as soon as we ascend to any mereological scale above that of fundamental physics, the sorts of properties, relations, and magnitudes we encounter cannot qualify as perfectly natural. (For clutter-reduction, I will also henceforth mostly speak just of "properties", even when relations and magnitudes are also intended).

Lewis offers a number of reasons, many persuasive and all intriguing, for thinking that we need to accept, as a fundamental metaphysical distinction, a distinction between perfectly natural properties and the rest. Set these reasons aside (you can find most of them in his 1983b). The question I wish to focus on, instead, is this: should we also take this distinction to be *graded*? That is, should we insist that, among all the properties that are *not* perfectly natural, some are nevertheless *more* natural than others? And if we do so insist, what sort of account can we give of what these gradations consist in?

Now, as to the first question, I'm going to assume that the answer is "yes". Lewis himself offers one very important reason that I'm going to set aside, which is that without such a graded distinction, it will be impossible to give a naturalistically acceptable account of the content of language and thought that does not face an insurmountable underdetermination problem (see the last section of his 1983b, as well as his 1984). I am, instead, go to lean upon a much more prosaic observation about scientific inquiry, which is just that, regardless of the mereological scale at which it operates, it seems to be a central and nonnegotiable feature of such inquiry that it aims to develop the right sorts of descriptive resources for describing the structure of the world at the given scale. There are, for example, ever so many ways that, logically speaking, one could describe reality at the scale at which chemistry operates. But the chemist's taxonomy is the best (or at the very least, one of the best); and I'm going to assume that to say that it is best is to say that this taxonomy tracks highly though not perfectly natural distinctions in nature.

So let us grant that we cannot do without a graded distinction between more and less natural properties. Then how do we answer the second question – how do we give an account of what these gradations consist in? That turns out, I think, to be an extremely difficult (and open) problem. For now, I would

like merely to consider and reject two approaches to it, as a way of highlighting how difficult it is.

First, one might simply treat gradations of naturalness as metaphysically primitive, and so unanalyzable. Superficially, this might seem an acceptable option; after all, Lewis himself argues that it is perfectly reasonable to accept the distinction between *perfectly* natural properties and the rest as primitive. But in fact I think this option is not acceptable. Here, briefly, are two reasons.

To begin with, we have granted that the pattern of instantiation of perfectly natural properties completely determines the qualitative structure of the world. And so, facts about comparative similarities and differences among objects that exist at scales above the most mereologically fundamental are fixed by this pattern of instantiation. But when, in the course of scientific (or even ordinary-life) investigation, we introduce distinctions among the less than perfectly natural properties, these distinctions earn their keep only insofar as they track explanatorily important similarities and differences among large-scale objects. So, if these explanatory distinctions are themselves ultimately grounded in facts about how the perfectly natural properties array themselves, then that just is to say that gradations in naturalness must be so grounded – and so cannot be metaphysically primitive, after all.

In addition, it's a good piece of philosophical methodology to avoid primitives that are ungainly. And in the present case, the imagined primitive seems *too* ungainly, since the way that gradations in naturalness are marked is quite complex. For example, for a chemist, the classification "being a methane molecule" will be more natural than the classification "being a methane molecule whose carbon atom is C-12": for whether the carbon atom is C-12 or C-13 will make no relevant *chemical* difference. But for a *nuclear* chemist, that difference *is* relevant, in a way that could reasonably reverse the judgment of naturalness. (It matters, for example, to the longevity of the methane molecule.) Cases like this suggest that the sort of naturalness that comes in degrees will exhibit a relativity to explanatory context: what count as the distinctions relevant to *some* explanatory projects may (even at the same mereological scale) differ from what count as the distinctions relevant to *distinct* explanatory projects. That complexity of conceptual structure makes it implausible that degrees of naturalness are simply metaphysically primitive, and at any rate deprives a philosophical account that treats them so of the resources it would need to explain this relativity to explanatory context.

So I'm going to henceforth assume that however attractive it may be to treat the category of *perfectly* natural property as primitive (in part, presumably, because the sort of relativity to explanatory context just discussed does not show up at the level of fundamental physics), an informative account is obligatory of the gradations in naturalness that the less than perfectly natural properties exhibit.

Lewis himself is perfectly aware of the need for an account of what makes one property *more natural* than another. He offers a simple and straightforward proposal, in the context of explaining how distinctions of naturalness yield distinctions in eligibility of reference:

> Indeed, physics discovers which things and classes are the most elite of all; but others are elite also, though to a lesser degree. The less elite are so because they are connected to the most elite by chains of definability. Long chains, by the time we reach the moderately elite classes of cats and pencils and puddles; but the chains required to reach the utterly ineligible would be far longer still. (Lewis 1984, p. 228)

Begin with a language whose non-logical vocabulary refers to *perfectly* natural properties; take this language to be rich enough that every perfectly natural property instantiated in our world gets referred to. Given some less-than-perfectly-natural property F, there will be *some* predicate of our canonical language – perhaps a very long, complicated predicate – that expresses it. F will be *more natural than* some other less-than-perfectly natural property G just in case the predicate expressing F is shorter than the predicate expressing G.

The proposal pretty clearly needs some refinement. *Many* predicates will express a given property; presumably we are to pick the shortest. While we can compare two predicates for length, we can also compare them for *simplicity*. Suppose the shortest predicate for F is slightly longer – but significantly simpler – than the shortest predicate for G;[6] should the advantage in simplicity outweigh the disadvantage in length? How shall we handle properties that are "multiply realizable", in the specific sense that predicates expressing them cannot be defined in a canonical language whose non-logical vocabulary refers only to *actual* perfectly natural properties? (For example, suppose the

[6] How would this happen? Well (for example), suppose the predicates contain quantifiers, and that those that appear in the predicate for F involve fewer alternations than those that appear in the predicate for G.

restriction of F to this world can be defined simply and efficiently, whereas F itself cannot; should we count F as natural-in-our-world?)

These are interesting questions, and perhaps even important, *if* our aim is to construct an account of the more/less natural distinction that will serve the purposes of a theory of the content of language and thought. But our aim here is different: it is to see how nature's ultimate joints – the joints given by the pattern of instantiation of perfectly natural properties, the joints it is the business of fundamental physics to discern – give rise to the less-than-ultimate joints whose study is the province of the special sciences. (Note well that I do *not* make Lewis's hasty assumption that a more/less natural distinction suited to *this* task will be one and the same as the more/less natural distinction suited to the theory of *content*.) And we do not need to resolve the foregoing questions to see that Lewis's proposal is on entirely the wrong track.

Consider the property of being methane. One early sign of trouble for Lewis's approach is that it looks as though, on that approach, it will be fixed once and for all whether this property is more natural than the property of being "stable" methane (that is, the property of being a methane molecule whose carbon atom is C-12), and that is just because the facts that Lewis's approach deems relevant – shortness and/or simplicity of definition in a canonical language – do not themselves exhibit any dependency upon explanatory context. There is, perhaps, the tiniest bit of wiggle room: maybe the trade-off between simplicity and length of definition could be taken to vary with explanatory context. But it would be pointless to pursue such a loophole, for the real, underlying trouble is much more straightforward: what this approach deems relevant to naturalness just bears no adequate connection to what underlies the explanatory utility of classifications in actual scientific practice.

A couple of examples should make the difficulty sufficient vivid. First, compare the property of *being methane* to the property of *being composed of 26 particles, each of which is a proton, neutron, or electron* (a property that most, but not all methane molecules share). Here is something indisputable: the shortest, simplest predicate in the canonical language expressing the first of these properties will be *vastly* longer and more complicated than the shortest, simplest predicate expressing the second. We *know* this: after all, I

just all but *gave* the canonical predicate expressing the second property[7], whereas, to produce the canonical predicate for the first property, one would need to begin with the standard chemist's definition of methane as "a molecule composed of four hydrogen atoms covalently bonded to a carbon atom", and proceed to unpack the predicates "is a hydrogen atom", "is a carbon atom", and "is covalently bonded to". And doing *that* will require bringing to bear substantial resources from theoretical chemistry and nuclear physics. It would take a while. You're rather unlikely to find anyone patient enough to be willing to pursue this project.

So, by the lights of Lewis's account of the less-than-perfectly-natural, the property of being composed of 26 protons, neutrons, and electrons ought to be *significantly* more natural than the property of being methane. And that's just silly. After all, methane molecules all have in common – in virtue of being methane molecules! – a wide variety of explanatorily important features. Thanks to the way that their covalent bonds affect their structure, they are all close to perfectly *tetrahedral*. They are all close to the same *size*. They all *react*, chemically, in exactly the same way. And so on. By contrast, *no explanatory purpose whatsoever* is served by distinguishing the class of things-composed-of-26-protons-neutrons-and-electrons. The vast majority of 26-pne's in the universe, after all, are *scattered*, their 26 different parts separated by light-years of space. And even those that are not exhibit no interesting or systematic behavior. (Except, of course, those that also happen to be methane molecules.) Just picture yourself writing a grant proposal – a *serious* one, mind you – asking for funding so that you can start a new program of research into 26-pne's.

Here is a second example, that reinforces the point that greater explanatory value of a classification can very often point in the *opposite direction* from greater simplicity or efficiency of canonical definition. Granted, *being methane* is a useful chemical property to know about. But *being a saturated hydrocarbon* is much *more* useful: it lends itself to a greater range of more important

[7] Assuming that protons, neutrons, and electrons are all fundamental particles, that is. Of course they're not (not protons and neutrons, anyway); but dropping this assumption would not make the slightest bit of difference to the plausibility of the claim that the canonical predicate expressing the second property will be vastly simpler and shorter than the canonical predicate expressing the first property.

generalizations, as any organic chemist will tell you.[8] So, measured by explanatory utility, the category of saturated hydrocarbons is a much more valuable one to distinguish than the category of methane molecules. But its canonical predicate will necessarily be longer and more complicated than the canonical methane-predicate. (Consider, for example, that this predicate must include a specification that the covalent bonds holding the molecule together are *single*, something that would be redundant in the methane-predicate.)

Examples like these convince me that there is something fundamentally misguided about Lewis's account of the less-than-perfectly-natural. It *may* be that brevity and/or simplicity of canonical definition plays some role in accounting for how nature's joint emerge at larger mereological scales. But it cannot be the whole story.

Of course, one could resist this conclusion – *if* one is willing to sever the close connection I have been taking for granted between "natural" and "explanatorily valuable". Perhaps that option could be profitably pursued; myself, I think it gives the game away. I will continue to take for granted that whatever else nature's joints are, they had better turn out to be distinctions that it is of the first explanatory importance to know about. And so, given the failure of Lewis's account of how large-scale joints are grounded in fundamental ones, it makes sense to turn to the theory of explanation itself for clues to an alternative.

3. Causal explanation and explanatory depth

Lewis's insight, summarized in the thesis Causal Explanation, is surely correct: to explain why an event occurs must involve giving information about its causal history. The reasons for this verdict are, I think, fairly obvious. For when we reflect on the abstract structure of our judgments concerning what causes what, and of our judgments concerning what explains what, we find that they are remarkably similar – too similar for this to possibly be a coincidence. In particular, in both domains we draw a firm distinction between events knowledge of which serves as a *good predictive basis* for other events, and events that *cause* or *explain* those other events; and we draw these distinctions in *exactly the same way*. For example, if Billy throws a rock at a window, and we

[8] Saturated hydrocarbons are molecules composed of hydrogen and carbon, where all chemical bonds are single covalent bonds (either carbon-carbon or carbon-hydrogen).

know that the rock is sufficiently hefty, very well aimed, and thrown with sufficient force, and that nothing stands poised to intercept it en route to the window, then we have an excellent basis for predicting that the window will break. But whether the throw explains the subsequent breaking, or is a cause of the subsequent breaking, depends on what is going on in the surrounding environment – and depends on those goings on in exactly same way. Thus, if Suzy *also* throws a rock at the window, and her rock gets there first, then Billy's throw neither explains nor causes this breaking, notwithstanding the fact that knowledge only of it would have allowed us to predict that the window would break.

These observations are, I take it, perfectly familiar, and it is largely because of them that causal theories of explanation are so dominant in the contemporary literature, having long since supplanted the logical empiricist deductive-nomological model, and having successfully resisted the incursions of various other accounts that downplay the importance of causation to explanation.[9]

All the same, there are legitimate and serious grounds for dissatisfaction, at least with Lewis's version of a causal theory of explanation. It will be instructive to highlight three, and then sketch some ways in which Lewis's account might be augmented and polished so as to deal with them.

First, the account remains far too schematic, without an account of what causation itself is. I say this, not out of some absurd notion that a philosophical theory that makes use of concept X thereby incurs an obligation to include an analysis of X; rather, there are reasons specific to causation that make this concept a poor choice of primitive, in an account of explanation. One is that causal relations between events at one scale are, very plausibly, *metaphysically grounded* in causal relations at smaller scales – and, ultimately, metaphysically grounded in the bare facts about the world's total history of complete physical states, together with the fundamental laws that dictate the evolution of those states. We would therefore deepen our understanding of what explanation is if we understood how this grounding works. More importantly, an account of explanation ought to make it clear why the acquisition of explanatory information is *valuable* for creatures like us – the sort of enterprise it makes

[9] I have in mind, for example, Kitcher's unificationist account (1989), and accounts that lean on probabilistic dependence – e.g., (Salmon, 1971).

sense to invest quite a lot of effort into. (This is a point that Jim Woodward has very effectively emphasized in his terrific recent work on explanation (Woodward, 2005).) And that sort of clarity will only be achieved via an account of causation itself. Finally, a review of the rich philosophical literature on causation will reveal that there are too many unanswered questions about causation for us to feel comfortable that we have an adequate grip on what the causal history of an event *is*. Is causation invariably transitive? Is causation by omission the same thing as ordinary causation? What about causation by double-prevention? Are the most basic causal relata really *events*, or should they rather be taken to be *facts*? All of these questions, and more, remain up in the air. (See (Hall & Paul, forthcoming) for extensive discussion.)

The second reason for dissatisfaction is that there is more – much more – to explanation than merely the explanation of particular events. In fact, in mature scientific inquiry, it is only very rarely that the explanation of particular events takes center stage. (E.g., a cosmologist might have as her life's work explaining the Big Bang; but you won't find many more examples like that.) Now, Lewis is perfectly aware of this fact, and makes no pretensions to having provided a complete philosophical account of explanation. Still, it is overwhelmingly plausible that the project of explaining particular events bears interesting connections to the other sorts of explanatory projects that scientists *do* put at center stage; and we shouldn't be satisfied with Lewis's account until it is developed in such a way as to make these connections clear.[10]

The third reason for dissatisfaction is that Lewis's account misses something of great importance to us when we seek the explanation of a particular event. An example will illustrate. A window has broken. Why? Because Suzy threw a rock at it. Now, we could obviously fill out that answer in many ways, thereby increasing the amount of explanatory information conveyed: we could trace the intermediate causes connecting Suzy's throw to the breaking; we could trace her throw's own causal origins; we could highlight the other causes contemporaneous with her throw with which it conspired in order to bring about the breaking. All of these ways of adding explanatory content Lewis's account, of course, recognizes. But it misses a distinct dimension along which our explanation of the window's breaking can be deepened. For what we might do instead is to highlight those aspects of Suzy's

[10] (Lewis, 1986b) includes some sketchy remarks on this topic. They do not suffice.

throw *in virtue of which it was able to bring about the breaking*, distinguishing them from other aspects that were causally (and so explanatorily) *irrelevant*. For example, the mass of her rock was important, but its color, not so much. And we could go further still: we could articulate, even with some mathematical precision, the structure of the way in which the window's breaking depended upon such factors as the rock's mass, the angle and velocity of the throw, and the distance between Suzy and the window. Granted, in any ordinary context all of this would be overkill – but that is not the point. The point here is that we have a kind of information that is clearly explanatory of the window's breaking, and that Lewis's account misses.[11] Again following Strevens (2009), I will say that this sort of information adds to an explanation's *depth*.

Let me now sketch an attractive approach – and one that has, I think, proved enormously philosophically fruitful over the last couple of decades – to understanding causation and explanation. It will go a fair way to answering the foregoing complaints. The central idea is extremely familiar: we should understand causation in terms of counterfactual dependence.

Now, this idea needs to be developed in the right sort of way, and two points in particular are critical to keep in mind. First, one should not think that the proper route to a counterfactual theory of causation is by way of some all-purpose semantic account of ordinary language counterfactuals (as Lewis himself apparently did; see his 1979). No, the counterfactuals in question need to be specialized. My own view is that they should have the following archetypal

[11] There is a complication, because Lewis in various places (for example 1986d) advocated a theory of events according to which at least some of the information being discussed here could be imported into the individuation conditions for the events themselves. Thus, he distinguishes events that are perfectly coincident in space and time on the basis of which of their features are essential, and which accidental. So perhaps you could say something like this: There were many throws that took place, of all of which Suzy was the agent; one of these was, inter alia, essentially a throw of a rock with such-and-such mass, but only accidentally a throw of a rock with such-and-such color. This throw caused the breaking. Other of the coincident throws – e.g., the throw that was essentially of such-and-such color a rock, but only accidentally a throw of a rock with such-and-such mass – did not. So you might hope – at least, if it's really important to you to preserve the exact letter of Lewis's account – that the sort of information that I am suggesting contributes to the explanatory depth of an event-explanation can simply be coded into the exact specification of the events that make up the target explanandum's causal history. I rather doubt this can be done, and I'm certain it cannot be done without producing a philosophical theory of events that is unpleasantly cumbersome. But at any rate, it doesn't really matter. The crucial point for the purposes of our discussion is simply that a good account of event explanation needs to recognize, somehow or other, the dimension of explanatory goodness that I've highlighted.

form: if the state of the world at time t had been just as it actually is, except with respect to goings on in a particular localized region, and if the state local to that region had differed in such-and-such a way, then the state of the world at a certain *other* place and time would have differed in such-and-such a way. There is a clear story to tell about how counterfactuals of this form are underwritten by the fundamental laws of nature; see for example (Maudlin, 2007b) or (Hall, 2011b). The basic idea is quite simple: given the alteration to the time-t state of the world specified in the antecedent, one simply updates the entire counterfactual history by plugging this state into the fundamental laws. Thus, this recipe shows how the fundamental laws, together of course with the totality of facts about our world's history, endows our world with a rich *localized dependence structure*.

The second point – which we will mostly set aside for the remainder of this essay – is that it will not do to simply *identify* causal structure with localized dependence structure. That is the lesson of cases of preemption, as for example the case mentioned earlier, in which Billy and Suzy both throw rocks at a window with deadly accuracy, but Suzy's rock gets there first. Here we see near-perfect symmetry between the relations of localized dependence holding between Suzy's throw and the breaking, and between Billy's throw and the breaking; but for all that, there is a striking asymmetry in causation. The right response, in my view, is not to try to exploit the tiny discrepancies in localized dependence structure that distinguish Suzy's throw from Billy's, but rather to recognize that part of what we're tracking when we track causal structure is the intrinsic structure of the processes that connect causes to effects. But this story gets quite complex, and at any rate, when told correctly, it still vindicates the thought that in ordinary cases, causal structure is relatively cleanly manifested in localized dependence structure. (Compare: we know better than to identify the property of *being disposed to Φ under conditions C* with the counterfactual property of being such that you would Φ if you were in conditions C; cases of masking and mimicking refute that simple equation. All the same, in ordinary cases the dispositional property is indeed manifested in the simple counterfactual behavior.)

So suppose that – while, again, bracketing the real and important worries raised by cases of causal preemption – we identify the sort of causal structure that is relevant to our explanatory interests with localized dependence structure. Then the sources of dissatisfaction mentioned above go away. We have an account of causal/explanatory structure that shows how this structure

is grounded in features of our world that are metaphysically more fundamental. Without going into details, the account can be used to say sensible things about thorny topics such as causation by omission, putative counterexamples to transitivity of causation, etc. (See (Hall & Paul, forthcoming).) We can sketch simple and attractive reasons why causal/exploratory information, so construed, is valuable to creatures like us: it is the sort of information in light of which we are able more effectively to navigate our world – a point emphasized by Woodward (2005) – and in addition it is the sort of information that creatures with our limited epistemic capabilities can reasonably hope to acquire, and by means of which we can reasonably hope to build up a more and more sophisticated understanding of the nomological structure of our world (Hall, 2011a).

We can also draw a connection between our prosaic practice of explaining ordinary events and the more refined and exalted explanatory aims of the sciences: whereas, in developing an explanation for why some particular event occurred, we are aiming to spell out one fairly *restricted* bit of the world's overall localized dependence structure, one of the central aims of the sciences is to discover *wide reaching* and *nomologically robust generalizations* concerning this structure, and patterns within it.

Finally, an account of explanatory depth falls out rather naturally. Consider our case of the broken window. By saying that it broke because Suzy threw a rock at it, we are conveying a bit of information about the localized dependence structure within which the breaking of the window is embedded. But we are doing so only very crudely: we're saying, roughly, that if the region of space in which her throw took place had differed just enough so that she didn't throw (but, say, stood idly by), then, given the laws, the entire state of the world would have evolved forward in such a way that the window did not break. We convey much more sophisticated information about the given localized dependence structure if, instead, we detail *which variations on Suzy's throw would or would not have led to a breaking*. Talk of the features of her throw in virtue of which it was a cause of the window's breaking is really just talk of the contours of such variations.

4. A puzzle about explanatory depth

But at this point, a puzzle emerges. Exploring it will take us directly back to the issues discussed in §1. To bring out the puzzle, we will draw again on our example of the broken window.

Here, in the abstract, is a way to see what we're doing when we increase the "depth" of our explanation of the window's breaking, by not merely citing the fact that Suzy threw a rock at it, but by articulating which counterfactual variations on her throw would and would not have led to a breaking. We have picked out a certain region of space-time: the region in which Suzy's throw takes place. Holding fixed the state of the rest of the world at the given time, the state of that particular region is, in the actual situation, such as to lawfully guarantee that the window breaks (at a certain time).[12] There are a multitude of nomologically possible alternatives to the exact physical state that this region instantiates. Some of these alternatives are such as to *still* lawfully guarantee the window's breaking (at roughly the same time, and, again, holding fixed the state of the rest of the world at the initial time); some are not. In aiming for explanatory depth, it appears that we are aiming to show how exactly the distinction between the former sorts of alternatives (example: an alternative in which the color of the rock is different) and the latter sorts (example: an alternative in which the rock is substantially lighter) is to be drawn.

But that can't be right, for we are doing some something more, and something much more subtle. Consider that one of these nomologically possible alternative states of the given region is the following: Suzy has no rock in her hand, but is in the process of running up to the window to level a vicious kick at it. Clearly, when we try to deepen our understanding of why the window broke by asking which sorts of variations on Suzy's throw would still have led to a breaking, we do not mean to include this scenario as one of them. Why isn't this alternative relevant, in the specific sense that it should be classed together, for explanatory purposes, with such alternatives as the one in which Suzy throws a rock of a slightly different color?

You might think the answer obvious: the actual cause of the window's breaking is a *throwing of a rock*, whereas whatever is going on in the imagined alternative – call it a "preparing to execute a running kick" – is a *different sort*

[12] Note that we are assuming determinism here. The story is more complicated if we relax that assumption, but not in ways that it would be profitable to explore.

of event altogether, and so cannot be seen as a "variation" on the actual cause. But to say that is to do nothing more than to highlight that we have already somehow managed to impose a certain scheme for taxonomizing events as the *explanatorily appropriate* one. It is, evidently, acceptable for our explanatory purposes to classify events as "throws" or as "preparations to kick", but not appropriate to use "disjunctive" classifications such as "throw or preparation to kick". (For if it were appropriate, then the counterfactual scenario in which Suzy is preparing to execute a running kick *would* count as a variation on the actual scenario; after all, she's *doing the same sort of thing*; just in a different way.) Why not? It seems to me we have not *answered* the original question, so much as forced it to take a different form.

Can we simply draw on the natural/non-natural distinction (really: the *more* natural/*less* natural distinction), at this point? Perhaps as follows: A way of classifying events that lumps together Suzy's actual throw with (inter alia) her counterfactual preparations-to-kick draws a much *less natural* distinction than a way of classifying that simply lumps together her actual and counterfactual throws; and it is *for that reason* that we achieve explanatory depth by deploying the latter classification, but not the former. Or, to put the point in terms of similarity, a counterfactual preparation-to-kick is *too dissimilar* to the actual throw, as compared to counterfactual variations on this throw, to count as one of the alternatives among which we need to distinguish, in order to achieve explanatory depth. (And these similarity facts, in turn, are grounded in the facts about the less-than-perfectly-natural properties instantiated in the actual and counterfactual scenarios.)

But this sort of appeal to the more/less natural distinction strikes me as far too cavalier. Given the problems raised in §1 for explaining what this distinction comes to, appealing to it doesn't *illuminate* so much as *label* what we are trying to understand. And at any rate, the presupposition that we achieve explanatory depth by focusing on those nomologically possible alternative states of the given region of spacetime that count as *variations on the actual throw* – aiming to distinguish those of them that lead to a breaking from those that do not – is false. Consider a variation in which Suzy throws the rock with slightly bad aim, just missing the window – but throws the rock so hard that it breaks the sound barrier, with the subsequent sonic boom shattering the window. We do not mean to classify this variation, *either*, together with variations in which we merely ring changes on the rock's color, etc. And this, notwithstanding that it is an alternative that clearly counts *as a*

variation on her throw. Again, why not? Not because lumping this variation together with the others produces an overly "unnatural" classification.

There *is* an answer, of course: breaking a window by way of a sonic boom counts, as compared with breaking it via direct impact, as a sufficiently different *way* of causing it to break, that our explanatory purposes are ill-served by lumping them together. That answer is correct, as far as it goes. But, again, it's a dead end to think that you can unpack that answer in some philosophically illuminating fashion by claiming that what *makes* our explanatory purposes ill-served in this way is that the classification in question is *insufficiently natural*.

And now for a radical suggestion: it's not just that this move leads to a dead end, it's that it gets things *exactly backwards*. What makes a classification that blends breakings-via-impact together with breakings-via-sonic-boom unnatural is that it ill-serves our explanatory purposes. In the remainder of this essay I am going to explore this idea.

5. Unification as a cognitive aim

Let's recap. A very good idea about causal explanation is that what we are seeking, when we set out to acquire such explanations, is information about particularly distinctive features of, or patterns in, the structure of localized counterfactual dependence that our world exhibits. One sort of structure in particular is the kind of structure knowledge of which gives our causal explanations "depth": it is the structure constituted by facts about how the localized state of the world in one place and time counterfactually covaries with the state in another place and time. Put another way, we are not merely interested in knowing that what goes on *here* and *now* causes what goes on *there* and *then*; we are also interested in tracking how this causal relation remains stable under counterfactual variations in the cause. But we are, it has emerged, not interested in just *any* old variations: for it appears to be bad explanatory policy to track variations in the cause under which its causal relation to the effect still obtains, but in an *overly different manner*. And so we have arrived at the need to understand what these distinctions among "ways of causing" are themselves grounded in.

What I have suggested is that it is a mistake to turn to the more/less natural distinction for help, because that distinction itself needs to be grounded in an account of what makes for better or worse satisfaction of our explanatory purposes. I will argue for that suggestion indirectly, by outlining a way that our

explanatory purposes could be served that could plausibly serve as grounds for a more/less natural distinction, and by leaving it a pregnant open question how *else* this distinction could be grounded.

It would be going in a circle, at this point, to give an account of our explanatory aims that merely returned to the themes discussed in the last section, asserting that these aims consist in the acquisition of causal information about the target explanandum. That is, of course, *one* of our aims, and a crucially important one at that; more generally, it seems to me that there is no way that we as scientific inquirers can come to an adequate understanding of our world, without knowledge of the metaphysical dependency relations (causal or otherwise) that knit it together. It's just that that cannot be the *whole* philosophical story about explanation. And what we need, at this point in the dialectic, is precisely the other part of the story. I therefore propose that one of our aims, in trying to develop an understanding of our world, is, in addition, to *develop cognitively effective means for organizing our information about the world*, in particular causal information.

That idea has clear connections to what, in the philosophical literature on explanation, has gone by the name of "unificationism" – which is, principally, Kitcher's unificationist account of explanation (1989). On this account, very roughly, explanations are arguments that instantiate very widely applicable patterns of argument. For reasons best left offstage, I do not think Kitcher's account succeeds, so let me hereby alert you that I do not in any way mean to be drawing upon it. I appropriate the label "unificationist" simply because, like Kitcher, I think that one important part of what we are after in explanation can be accurately (if very incompletely) described as the acquisition of a *unifying picture of the world*.

Now, the idea that one of the central things we are after in explanation is the development of cognitively effective means for organizing our information stands desperately in need of development itself. I do not have a theory to offer of just what a "cognitively effective means of organizing" is (and not, alas, merely for reasons of lack of space). But it is easy enough to find evocative examples that, I think, do an extremely effective job of bringing out the unificationist strand in our thinking about explanation. Here is one that is slightly goofy, but for all that one of my favorites.

Consider the following initial segment of an infinite sequence of natural numbers:

1,1,1,2,3,2,1,3,5,4,2,5,7,8,3,7,9,16,5,11,11,32,8,13,13,64,13,17,...

Perhaps you've figured out the rule that generates the sequence. Perhaps, on the other hand, you find it confusing. You don't *understand* it. You don't know *why it has the form it does*. If so, the following way of reorganizing the initial segment will make things crystal clear:

1,	1,	1,	2,
3,	2,	1,	3,
5,	4,	2,	5,
7,	8,	3,	7,
9,	16,	5,	11,
11,	32,	8,	13,
13,	64,	13,	17, ...

Looking down the columns, we see that the sequence is just an interleaving of the odd numbers, powers of 2, fibonacci numbers, and prime numbers. Once you see this, you understand the sequence. But not by acquiring a special sort of information about it. (The sequence is, after all, not the sort of thing that has "causes", or that "metaphysically depends" on anything else.) To me, examples like this evoke in its purest form the idea that to understand some subject matter is to organize one's information about it in the right sort of way.

Not surprisingly, examples with this particularly clear character – in which explanatory insight is achieved not at all via the provision of a special sort of dependency information, but entirely by organizing the information we have in the right sort of way – are much easier to find in mathematics than in the sciences, simply because in mathematics the only kind of dependency information that's available is information about logical entailment, and that only gets you so far, explanatorily speaking. Just consider the fact that mathematicians routinely distinguish proofs that are illuminating from proofs that aren't; and yet the unilluminating proofs are, for all that, proofs! So something else must ground the distinction. I suggest that the something else concerns how the illuminating proofs generalize to other results, how they highlight easily overlooked connections between their subject matter and other mathematical topics, and so on; in short, they are illuminating to the extent that they contribute to the effective organization of mathematical knowledge.

In the sciences, by contrast, explanation almost always involves the provision of interesting, distinctive dependency information, and for that reason it can be difficult to see that unificationist requirements on understanding also play an important role. Still, some examples bring out these requirements rather nicely. Consider the periodic table of the elements, which

is unquestionably of immense value to us in enhancing our understanding of the chemical and atomic behavior of atoms, and which has this value *precisely* because of the brilliantly effective way in which it organizes our information about this behavior.

So let's grant that explanation even in the empirical sciences involves both distinctively metaphysical and distinctively psychological aspects: on the one hand, we want a special sort of information – information about what depends on what, metaphysically speaking – but on the other hand, we want our information, especially our dependency information, to be organized in the right sort of way – where what makes for good organization, presumably, depends on potentially quite idiosyncratic features of human psychology.[13] Then the next thing to notice is that what makes for good organization is very often going to be a holistic matter.

Consider, again, the periodic table of the elements. To be sure, what makes this such a powerfully effective tool for understanding is in part that is a table of the *elements*: and the distinction between elements and non-elements does not obviously involve any holistic considerations. (Rather, it seems that we focus, for explanatory purposes, on elements simply because they are highly stable configurations of matter, and so the sorts of things about which it is possible to make useful generalizations concerning their behavior.) But it is the *periodic* table of the elements because of the way in which it classifies elements into different chemical types. And what makes the particular scheme of classification built into the table so explanatorily superior to the multitude of logically possible rivals cannot, I think, be appreciated by examining its components piecemeal. It is not, as it were, that a certain amount of explanatory goodness attaches to any scheme that distinguishes noble gases from things that are not noble gases, and a certain additional amount of goodness attaches to any scheme that incorporates a distinction between metals and non-metals, and so on; with the overall goodness of our own scheme simply being the sum of these individual goodnesses. No, it is because of the way in which our scheme *as a whole* arranges our knowledge of the chemical and atomic features of the elements that it is so explanatorily powerful.

[13] Well, maybe not. Maybe, indeed, the very possibility of rational thought requires that understanding be achieved partly by the imposition of a priori principles of organization. If you're obsessively concerned to preserve the pure, unadulterated objectivity of our explanatorily valuable classifications, that might be the way to go.

I think that something very similar happens – albeit in a fashion that is less rigorous, and much more difficult to discern – in our thinking about causation and causal processes. We saw above that we distinguish a breaking-via-sonic-boom as a sufficiently different way of breaking a window from a breaking-via-impact that it would be a bad mistake, given our explanatory aims, to count the latter sort of breaking as a variation on the former. It is enormously philosophically tempting to think that this distinction must be grounded entirely in features specific to the two kinds of breaking: just by, as it were, closely inspecting paradigm examples of the two kinds of breaking, one would be able to see that our explanatory aims require us to distinguish them *as* separate kinds.

But I think that is a mistake. I think it is much more plausible that in coming to grips with the vast profusion of causal processes we encounter even in ordinary life, we very early on (and almost certainly unconsciously) hit upon certain schemes for organizing these processes into types. Now, to borrow an idea from Lewis's work on laws of nature – see (Lewis, 1983b) and also (Loewer, 1996) – two extremely important desiderata we impose on candidate schemes (again, not consciously!) are, plausibly, the following:[14] First, it's good for a candidate scheme to be *simple*, not necessarily in the sense that it includes a small number of types, but perhaps in the sense that it makes use of a small number of basic parameters to characterize those types. Second, it's good for a candidate scheme to have the resources needed to express *powerful, informative* generalizations about causal structure (in particular, the sorts of generalizations that populate the special sciences, and that in the philosophical literature typically go, misleadingly, by the name "ceteris paribus laws"). These desiderata work in tension: consider that one way to get a simple scheme is to let the sole type of event be "event" and the sole type of causal process to be "causal process". Whatever you said about the causal structure of the world, by means of this scheme, could be said quite simply. But not very informatively. It is immensely plausible that achieving the best balance between these desiderata will involve holistic considerations.

Now, maybe it's hubris to think that the schemes we humans have developed are the best possible, for purposes of effective organization.

[14] Not the only two, surely. For example, it's plausibly a desideratum that a scheme for taxonomizing events and causal processes not yield up kinds whose membership is difficult or impossible for creatures like us to empirically determine.

(Certainly, one of the striking effects that a good scientific education can have on one is opening one's eyes to the availability of very different schemes, that sometimes improve dramatically on our ordinary ones.) But I will suppose that they are good enough. At any rate, what I would like to suggest is that lumping the two kinds of breakings together is an explanatorily bad move *not* solely because of features intrinsic to paradigm instances of each, but because no scheme for organizing the vast amount of information we possess about the causal structure of our world that did so could possibly meet the desiderata on effective organization as well as the scheme we have arrived at.

6. Conclusion: mapping the large-scale joints of the world

Let's return now to the questions left over from section 1. How is it that the natural/non-natural distinction, as it appears at scales above the most mereologically fundamental, is determined by the fundamental physical structure of the world (the structure given by the pattern of instantiation of perfectly natural properties, together with the fundamental laws of nature)? How does the structure of reality at the most fundamental levels determine the map of reality's joints at less fundamental scales? The right answer, I think, is that it doesn't – at least, not alone. The picture is rather the following: given how the world is fundamentally (where I take this to include: how its fundamental laws are), the world has a perfectly definite localized dependence structure, which for purposes of keeping things simple (i.e., ignoring the complexities that cases of preemption introduce) we will take to just *be* its causal structure. But this structure does not, as it were, come equipped with a uniquely best way to describe it, even at a given scale. Rather, we impose on it various taxonomies – different ones for different scales, certainly, and sometimes even different ones at the same scale, given that it can sometimes be useful for us to highlight certain patterns in the world's localized dependence structure at the expense of others. These taxonomies sort events and the causal processes that knit them together into kinds, and do so subject to the constraint that the sorting provide us with maximally effective tools for organizing our view of the causal structure of the world at the given scale. The map of the large-scale joints of the world is just constituted by whatever distinctions figure in such optimal taxonomic schemes.

It follows that, in a certain sense, the distinction between more and less natural properties at larger mereological scales fails to be perfectly objective:

for this distinction is determined in part by which taxonomic schemes do the best job for creatures like us of providing tools for the efficient and effective representation of causal structure. So it would be a bad mistake to think that what we are doing when we investigate the large-scale structure of the world is merely discovering the natural distinctions that are there to be drawn. But it would be just as bad a mistake to think that how to draw these distinctions is somehow entirely up to us. To say that is simply to forget that the fundamental physical structure of the world – which, I'm supposing, is what it is quite independently of facts about the structure of human cognition – is also an indispensable ingredient. In sum: It is a complex interplay between purely objective facts about reality's physical structure, on the one hand, and psychological facts about the structure of human cognition, on the other, that grounds the "joints" that nature exhibits at large scales.

Now, just in case this point wasn't obvious, what I am offering is *not* a proper theory of the more/less natural distinction. What *Lewis* offered, in his proposal that the naturalness of a property is fixed by the length of the shortest canonical predicate expressing it, was a proper theory. (Granted: it slips from "proper theory" back to "approach" if we amend it by saying that simplicity of the predicate also matters, while leaving it vague how simplicity itself is to be measured, and how simplicity and length trade off.) What I have offered are remarks that point in the direction of a theory. My hope is that they point, at least, in the *right* direction. At any rate, they pretty clearly cry out for elaboration.

Three avenues in particular seem to me worth pursuing. First, the picture I've sketched needs input from empirical psychology, since that is where we can hope for insight into how it is that organizing schemes in fact function in human cognition. Second, it would be helpful to explore how our taxonomizing strategies work when applied to toy models – Conway's game of "life", say.[15] Third (and relatedly), it would be helpful to explore case studies from especially well-developed and mature special sciences – organic chemistry, say.

It's highly unlikely that the results of such inquiry will yield anything as pristine as Lewis's account. No, it's going to be messy – and, maybe, messy in case-specific ways. For example, the way the natural/nonnatural distinction

[15] See for example http://en.wikipedia.org/wiki/Conway's_Game_of_Life.

plays out in organic chemistry may not be the same as the way it plays out in, say, evolutionary biology. But that's to be expected, if indeed this distinction results from an interplay between facts about physical structure and facts about human cognition in the way I have suggested.

REFERENCES

Hall, N. (2010). David Lewis's Metaphysics, *Stanford On-Line Encyclopedia of Philosophy.*

Hall, N. (2011a). Physical and metaphysical possibility. Manuscript.

Hall, N. (2011b). Causation and the sciences. In S. French & J. Saatsi (Eds.), *Continuum Companion to the Philosophy of Science*. London: Continuum, 96–119.

Hall, N. & Paul, L.A. (forthcoming). *Causation and the counterexamples: A user's guide.* Oxford University Press.

Kitcher, P. (1989). Explanatory Unification and the Causal Structure of the World. In P. Kitcher & W. Salmon (Eds.), *Scientific Explanation.* Minneapolis: University of Minnesota Press, 410–505.

Lewis, D. (1979). Counterfactual Dependence and Time's Arrow. *Noûs*, *13*, 455–476. Reprinted with Postscripts in (Lewis, 1986a), 32–66.

Lewis, D. (1983a). *Philosophical Papers, Volume I.* Oxford: Oxford University Press.

Lewis, D. (1983b). New Work for a Theory of Universals. *Australasian Journal of Philosophy*, *61*, 343–377. Reprinted in (Lewis 1999), 8–55. Page references are to this latter printing.

Lewis, D. (1984). Putnam's paradox. *Australasian Journal of Philosophy*, *62*, 221–236.

Lewis, D. (1986a). *Philosophical Papers, Volume II.* Oxford: Oxford University Press.

Lewis, D. (1986b). Causal Explanation. In (Lewis, 1986a), 214–240.

Lewis, D. (1986c). Events. In (Lewis, 1986a), 241–269.

Lewis, D. (1994). Humean Supervenience Debugged. *Mind*, *103*, 473–490.

Lewis, D. (1999). *Papers in Metaphysics and Epistemology*. Cambridge: Cambridge University Press.

Lewis, D. & Lewis, S. (1970). Holes. *Australasian Journal of Philosophy*, *48*, 206–212. Reprinted in (Lewis 1983a), 3–9.

Loewer, B. (1996). Humean Supervenience. *Philosophical Topics*, *24*, 101–127.

Maudlin, T. (2007a). *The Metaphysics Within Physics*. Oxford: Oxford University Press.

Maudlin, T. (2007b). A Modest Proposal Concerning Laws, Counterfactuals, and Explanation. In (Maudlin 2007a).

Salmon, W. (1971). Statistical Explanation. In W. Salmon (Ed.), *Statistical Explanation and Satistical Relevance*. Pittsburgh: University of Pittsburgh Press: 29–87.

Sider, T. (2011). *Writing the Book of the World*. Oxford: Oxford University Press.

Strevens, M. (2009). *Depth: An Account of Scientific Explanation*, Cambridge, MA: Harvard University Press.

Woodward, J. (2005). *Making Things Happen: A Theory of Causal Explanation*, Oxford: Oxford University Press.