DOI: 10.1111/nous.12433

ARTICLE

NOÛS

Naturalness by Law

Verónica Gómez Sánchez

New York University

Correspondence
Verónica Gómez Sánchez, New York University
Email: veronica.gomez@rutgers.edu

1 | INTRODUCTION

Meaningful predicates come in two kinds. Predicates of the first kind characterize ways in which objects can resemble each other other; two examples are ‘electron’ and ‘red’. Predicates of the second kind don’t correspond to any real dimension of similarity; two examples are ‘electron or red’ and ‘such that something is red’. Underlying this distinction between predicates is a distinction in reality: predicates of the first kind express natural properties and predicates of the latter express unnatural or gerrymandered properties.[1]

I follow David Armstrong (1978), David Lewis (1983), Ted Sider (2011), and many others in thinking that naturalness is not an arbitrary linguistic projection, but a useful metaphysical posit with a rich theoretical role. Only natural properties make for objective similarity among objects, figure in laws and other projectible generalizations, normatively constrain inductive inference, are involved in causal relations, are mentioned in good explanations, and are reference magnets (that is, highly eligible as meanings of our primitive concepts).[2]

If naturalness is to play this rich theoretical role, it must span many levels of description. We need a notion broad enough to apply not only to fundamental physical properties like charges or masses, but also to non-fundamental properties like *green*, *kangaroo* or *pain*.[3] This paper develops and defends a reductive account of a suitably broad notion of naturalness: for a property to be natural, I argue, is just for it to figure in a law of nature.

My defense of this account proceeds as follows. I first argue that naturalness, in the broad sense, is not metaphysically primitive; naturalness facts must be rooted in the world’s fundamental physical structure. Next, I consider a challenge for this idea, which appeals to the possibility of multiply realizable natural properties. I then develop a nomic account of naturalness that can meet this challenge, while also explaining how naturalness facts derive from the fundamental physical structure. Finally, I argue that the notion of naturalness yielded by my account is well-suited to play many of the key theoretical roles that the notion was introduced to play in the first

[1] I assume an abundant conception of properties throughout this paper. Setting aside complications having to do with vagueness, we can assume that each meaningful predicate expresses an abundant property. I use italicized predicates as names for the corresponding abundant properties.

[2] I discuss some of the connections between naturalness and other notions in §4. See Hawthorne and Dorr (2013) for comprehensive description of these and other roles that have been attributed to the notion of naturalness.

[3] Schaffer (2014) emphasizes this point.

place, and that those roles which it can't play are better played by the closely related notion of metaphysical fundamentality.

2 | NATURALNESS AND FUNDAMENTALITY

2.1 | Against Primitivism

It can be tempting to think that we do not need (and should not hope for) a reductive account of naturalness. Perhaps truths involving naturalness characterize the basic structure of reality, and there are no further facts in virtue of which this structure is in place. Call the view that all naturalness facts are metaphysically fundamental 'primitivism' (Schaffer, 2004).

Primitivism is most naturally regimented in terms of a second-order predicate $Natural$. It states that truths of the form 'Predicate 'F' is natural' are underwritten by truths of the form $Natural$(F), where the latter truths are not about the structure of the language we happen to use to describe reality, but rather about how reality itself is structured.

Some philosophers would reject this view on the grounds that primitive naturalness facts would be too spooky or unknowable: 'metaphysical' as opposed to 'empirical'. This is not the kind of complaint that motivates my account. Like many other naturalness enthusiasts, I am skeptical of this dichotomy between the empirical and the metaphysical. In fact, in giving my account, I too will need to help myself to some primitive notions that some would deem objectionably 'metaphysical', like the notion of a fundamental property.[4]

My concern with primitivism is specific to non-fundamental natural properties. To account for natural properties in non-fundamental reality, the primitivist must recognize that, for some non-fundamental F, $Natural$(F) holds fundamentally, which is unappealing for a number of reasons.

For one thing, the possibility of such fundamental facts is ruled out by an attractive principle about fundamentality defended by Ted Sider (2011):

> PURITY
> No fundamental truth involves a non-fundamental property.

PURITY connects two different notions of fundamentality: a sentential notion that applies to propositions, and a sub-sentential notion that applies to properties. The sentential notion of fundamentality applies to true propositions that hold in virtue of nothing else; truths that are stopping points for metaphysical explanation. The sub-sentential notion is meant to apply, for instance, to masses, charges, topological and geometric properties of space-time…, or to whatever properties would come to replace these in the ultimate fundamental theories for our world.

[4] Shamik Dasgupta (2018) has recently articulated another important challenge for primitivism (which applies just as well to primitivism about fundamentality). As I understand Dasgupta's arguments, what they threaten is the idea that a primitive distinction between properties could be 'intrinsically normative' or 'primitively normative'—normatively significant by its nature alone, independently of any facts about us. It is hard to say to what extent people like Lewis and Sider need to be committed to this idea. In any case, I am fairly certain that the ideas in this paper do not depend on the claim that the world's basic structure has this sort of intrinsic normative significance. I expect the grounds for the normative significance of naturalness to mention something about us: the fact that we are agents, that we value truth, that we have certain goals, that we are subject to certain epistemic constraints, etc. If this renders my notion of naturalness 'parochial' or insufficiently 'objective', then so be it. To my mind, this is only a problem if it undermines the authority of plausible normative principles invoking naturalness or fundamentality (e.g., relating to induction and/or good explanation), but I am not convinced that it does.

There is a straightforward argument from PURITY to the negation of primitivism. Take a non-fundamental property that is plausibly natural, such as *green*. To account for the naturalness of green, the primitivist has to say that $Natural$ (green) expresses a fundamental truth. But this is ruled out by PURITY because the truth in question would involve *green*, a non-fundamental property.

One limitation of this argument is that it invokes the potentially murky notion of involvement. If we think of propositions as having constituent structure, much as sentences do, involvement is easily definable: a proposition involves a notion if that notion is one of its constituents. But the structured conception of propositions is controversial, and without it, the appeal to an intuitive notion of involvement may be misguided.

For those who are skeptical of structured propositions, I offer a different argument against primitivism, which relies on another attractive principle connecting sentential and sub-sentential fundamentality:

> ATOMISM
> If p is a fundamental truth, then p is expressible either by a logically simple sentence or the negation of a logically sentence simple in a 'fundamental language'.

A fundamental language (Sider, 2011) is an interpreted formal language whose basic predicates express fundamental properties/relations, and whose individual constants all denote fundamental entities. I will assume, for the purposes of this paper, that such a language has the syntax and logical resources of a standard higher-order language.

ATOMISM is motivated by two ideas: (i) that truths which only admit of logically complex translations into a fundamental language are metaphysically explained by (or grounded in) logically simpler truths, and (ii) that fundamental truths are those that lack metaphysical explanations (or grounds). Insofar as we are able to explain/ground conjunctive truths in terms of their conjuncts, disjunctive truths in terms of their disjuncts, and quantifiactional truths in terms of their instances, we should regard them as non-fundamental (Fine 2012).[5]

We can argue from ATOMISM to the conclusion that $Natural$(green) expresses a non-fundamental truth. Since *green* is non-fundamental, it seems like the only way to express the truth that *green* is natural in fundamental vocabulary would be: $Natural(\lambda x\phi(x))$, where $\lambda x\phi(x)$ is a logically complex definition of green in fundamental terms. $Natural(\lambda x\phi(x))$ will be neither logically simple, nor the negation of a logically simple statement. It follows, by ATOMISM, that *green*'s being natural is non-fundamental.

There is a weightier reason to dislike primitivism, which does not rely on potentially controversial principles about fundamentality: primitivism fails to capture certain explanatory connections between naturalness facts and fundamentality facts.

One of the key roles that the notion of naturalness is supposed to play is explaining objective similarities among objects. If two objects share a natural property, they are thereby similar in

[5] A complication: suppose that $p\&p = p = p \vee p$ for any p. Then it would be wrong to think, as the paragraph above might be taken to suggest, that any truth which is equivalent to a conjunction or a disjunction in a fundamental language has a metaphysical explanation. To avoid this problematic implication, by 'conjunctive truth' we must understand a truth which is most perspicuously stated conjunctively—where perspicuity could be cashed out in terms of complexity in a fundamental language. The same goes for 'disjunctive truth' and 'quantificational truth'. The claim above is, then, that whenever the most perspicuous translation of a truth p to fundamental vocabulary is logically complex, that truth can be explained by its conjuncts, disjuncts or instances.

some respect; the more properties they share, the more they resemble each other overall. Now consider the property expressed by 'water molecule'. All the entities that instantiate this property are similar in some respect. And this similarity is not coincidental, but explained by the fact that they are all water molecules. It seems, however, that the similarity between these objects is also explained by the fact that they share a common chemical structure. Presumably the predicates 'hydrogen', 'oxygen', 'bonded', and 'parthood' all express natural properties/relations. So the fact that all water molecules have as parts two hydrogen atoms bonded to an oxygen atom seems to be enough to explain the respect in which they are similar.

To capture both of these explanations, the primitivist must posit two distinct explanatory paths: one going from the naturalness of *water molecule* to the similarity of its instances, and one going from the naturalness of *hydrogen*, *oxygen*, *bonded* and *parthood* (together with the chemical definition of *water molecule*) to that same similarity fact. But these two explanatory paths would be entirely disconnected: if *hydrogen* and *oxygen* hadn't been natural, water molecules would still have been similar merely in virtue of being water molecules. The same point can be made without the counterpossibles: what seems to be a single respect of similarity among water molecules is getting fully grounded twice over—once at the atomic level, and once again, independently, at the molecular level. The primitivist then has more structure than is needed to explain the facts; her view involves unparsimonious redundancy.

2.2 | The fundamentality-first approach

The three arguments above reveal that primitivism distorts the relation between naturalness and fundamentality. A promising idea for solving these problems is to take sub-sentential fundamentality as primitive, and to try to recover naturalness facts involving non-fundamental properties from fundamentality facts.

This 'fundamentality-first approach' has been at the center of the literature on naturalness: its proponents include David Lewis and Ted Sider. This approach grants that, where X is a fundamental property/relation, the question 'why is X fundamental?' has no answer. Thus, if spatio-temporal distance is a fundamental relation, this fact has no metaphysical explanation.[6]

The main challenge for this approach has to do with multiply realizable properties. Consider the property of being in pain. It is standardly assumed that this property can be instantiated by organisms with entirely different physical make-ups: humans, chickens, octopuses, aliens, and—perhaps one day—by non-carbon-based AIs. In view of this, what reason is there to suppose that similarity with respect to pain can be explained in terms of common micro-physical structure?

The suspicion that physical structure won't be enough to support the naturalness of multiply realizable properties has yet to be dispelled. Only one proposal for getting naturalness facts from physical structure exists to this day, and multiple realizability is its Achilles' heel.

The proposal in question comes from Lewis (1986). He suggested that the degree of naturalness of a non-fundamental property depends on how closely connected it is to the fundamental properties, where proximity is measured by the simplicity of the simplest definition of the non-fundamental property in terms of the fundamental ones. More precisely, a property's degree of naturalness is the multiplicative inverse of that property's 'FL-complexity', which is the

[6] As before, we can express these basic truths by means of a higher order predicate Fun, which combines with a first-order predicate to yield a proposition. A truth of the form $Fun(F)$ would say (roughly) that F expresses a fundamental property (without carrying any ontological commitment to F).

complexity of the simplest open sentence which expresses the property in question in a fundamental language, and contains no individual constants (which plausibly suffices for expressing a qualitative property).

There is much to like about this account. It gives us a simple and precise reductive account of naturalness, and it re-establishes the explanatory connections between naturalness facts and the world's fundamental structure, thus overcoming the problems for primitivism mentioned above. Unfortunately, this proposal makes the wrong predictions about the degrees of naturalness of multiply realizable properties, as a quick argument will show.

What might an FL-definition of *pain* look like, given that it is multiply realizable? Let a 'pain-realizer' be a property that has an FL-definition, and whose possession by someone fully grounds their being in pain. We might attempt to define *pain* by constructing a disjunctive formula in FL, where each disjunct defines one pain-realizer and all pain-realizers are defined by some disjunct. The resulting formula would not be an adequate definition of *pain*, if this property can be realized by 'alien' properties—i.e., properties that are not around in our world. But, even setting aside alien realizers, Lewis's proposal runs into trouble.

As others have pointed out (Hawthorne and Dorr (2013), Schaffer (2013)), this sort of disjunctive definition could turn out to be infinitely complex. And, even if it turns out to be finite, the account still makes incorrect predictions. Suppose that the simplest FL-definition of *pain* is a finite disjunction of pain-realizer definitions $\phi_1, \phi_2 \dots \phi_n$. The disjunction of definitions $\phi_2 \dots \phi_{n-1}$ will be syntactically simpler, but the property it defines is clearly less natural than *pain*—it is not projectible, explanatory, easy to refer to…

Sider (2011) suggests an alternative strategy for defining properties like pain in FL. Schematically: to be in pain is to have some property or other that is causally/lawfully connected in such and such way to stmuli $S_1, S_2 \ldots S_n$ and to behaviors $B_1, B_2 \ldots B_m$. Even granting that we can re-express this sort of definition in purely fundamental terms by substituting the 'S-terms', the 'B-terms', and the relevant causal or nomic vocabulary with adequate FL-definitions, we can show that FL-simplicity fails to correlate with naturalness. Take the property of having some property or other that is causally/lawfully connected in the relevant ways to stimuli $S_1, S_2 \ldots S_{n-1}$ and behaviors B_1, $B_2 \ldots B_{m-1}$. This property will have a simpler functional definition than *pain*, but this does not make it more natural.

2.3 | A dilemma

A few pages ago, thinking about properties like *water molecule* led us to suspect that non-fundamental properties inherit their naturalness from the way they are grounded in fundamental properties. But thinking about properties like *pain* complicates this picture, for we cannot easily trace any direct explanatory link between the way *pain* is grounded and its high degree of naturalness.

A potential dilemma now comes into focus. If naturalness facts are primitive—and so detached from the world's fundamental structure—there will be no room for explanatory connections between the naturalness of *water molecule* and its ultimate physical basis. If, on the other hand, facts about higher-level similarity are fully explained by the fundamental physical structure of the world, how do we get multiple realizability?

A hybrid view escapes this dilemma: the view that some naturalness facts are fundamental (those concerning multiply realizable properties), while other naturalness facts are grounded. But this hybrid view is inelegant. Having granted that we can recover much of the structure of natural domains like chemistry and biology from physical structure, is it sensible to insist that new fundamental joints in nature are to be found once we reach the level of minds? (Moreover, this hybrid view fares no better than primitivism with respect to the first two arguments in §1.1.)

The rest of the paper develops an account of naturalness that avoids the dilemma, without taking the naturalness of certain multiply realizable properties as brute. My account draws on some elements of Lewis's fundamentality-first approach: it presupposes a sub-sentential notion of fundamentality, and appeals to chains of definitions connecting properties across levels. However, the account does not take naturalness structure to reduce to fundamentality structure alone. In order to understand how certain multiply realizable properties can be natural, we need to take into account their privileged place in our world's nomic structure.

3 | THE NOMIC ACCOUNT OF NATURALNESS

The idea developed in the rest of the paper is simple: for a property to be natural is for it to figure in a law of nature. Despite its intuitive appeal, this nomic account of naturalness has rarely been explicitly defended in the literature.[7] The reason, I suspect, is that the nomic structure that it presupposes is unfamiliar and may seem problematic. In what follows I begin to characterize the notion of lawhood that the nomic account requires, and outline some challenges for it that the following section will address.

3.1 | Properties in laws

The nomic account requires that we make sense of involvement or constituency relations between laws and properties. Believers in Russellian structured propositions—abstract entities that have their truth conditions essentially and have properties as constituents—will have no trouble with this, if they take laws to be special kinds of propositions. But the nomic account does not depend on this structured conception of propositions. If we think of laws as special kinds of (interpreted) sentences, we can say that to be natural is to be expressed by a simple predicate that is a constituent of some law.

The idea that laws are sentences may seem unappealing at first. If we only have the sentences that belong to languages currently in use, then the view that laws are sentences rules out the possibility of laws that no language in use can express, and what laws there are becomes problematically dependent on linguistic practice. But this worry is avoided if laws are sentence-like set-theoretic entities that exist necessarily (e.g., sequences of pure sets). Unlike propositions, these abstract entities are not inherently meaningful, and so cannot be true or false simpliciter. We can nonetheless speak of them as being true under an interpretation, and reformulate the nomic account as follows:

$Natural(F) =_{def}$ there is a (possibly abstract) sentence-like structure s and interpretation I such that:

(i) s is a law under I, and

[7] Loewer (2021) proposes a similar nomic account of the notion of a fundamental property. My account, unlike his, will presuppose this notion of fundamentality and use it to account for high-level laws and high-level naturalness.

(ii) I maps some predicate constant in s to F.[8]

When I say of an English sentence that it is a law, this should be read as shorthand for a more complex claim: that there is some abstract sentence s which is a law under some interpretation I, and captures (so interpreted) roughly the same information as the English sentence in roughly the same way, that is, by means of semantically similar constituents arranged in syntactically similar ways.

(In what follows I will tend to drop the interpretation function argument in the term law, and I will speak as if each sentence comes uniquely interpreted. This simplifies some of the definitions to come, but everything I say can easily be paraphrased so as to make the appeal to interpretation functions explicit.)

3.2 | Permissive Lawhood

If we want to say that *spinach*, *person*, and *city* are natural properties, then—according to the nomic account—we need to posit laws involving each of them. This might seem odd if one is thinking of laws as the kinds of simple equations that fundamental physics aims to formulate, with paradigms such as Newton's laws or Schrodinger's equation.

To see the appeal of the nomic account, we must first conceive of a notion of lawhood that is much more permissive than the notion of a fundamental physical law. This permissive notion is, I think, the notion of lawhood that is relevant to the 'special sciences': sciences that predict and explain phenomena in non-fundamental terms. Laws, in the sense that interests me, are simply those generalizations that capture real patterns in nature, and play a special role in our understanding of the world—we rely on them for prediction, explanation and hypothetical/counterfactual reasoning.[9]

Given a suitably permissive conception of lawhood, some statements about spinach may well be laws. For example, 'Spinach is iron-rich' captures a widespread regularity in our environment that we can rely upon to predict what will actually happen, and to decide what would happen under hypothetical circumstances. Moreover, this regularity underwrites the explanatory link between pairs of propositions such as: *I eat more spinach in 2020*, *my iron levels increase in 2020*. The same goes for *person* and *city*: it may well be that the the most powerful scientific theories that one can state about the social would involve these properties.

Laws in the permissive sense differ in important respects from fundamental physical laws. The most joint-carving notion of natural necessity may fail to apply to many of them.[10] Moreover, fundamental physical laws are standardly thought to hold universally,[11] whereas many generalizations that I will be calling 'laws' have exceptions. For these reasons, they are often called 'robust regularities' or 'lawlike regularities' rather than 'laws'. (I have also called them 'crystallized regularities'.) Any of these terms could be used to state the account of naturalness defended here.

[8] For simplicity, I will help myself to the assumption that there are interpretation functions in the first-order domain; the desired definition could also be stated by means of higher-order quantifiers.

[9] I discuss this notion of lawhood and its theoretical role in more detail in Gómez Sánchez (2020).

[10] See Strevens (2008b) and Mitchell (2002) for arguments that the statements that play the role of laws in various special sciences are physically contingent.

[11] But see Cartwright (1983).

3.3 Fine-Grained Lawhood

While the notion of law that I'm after needs to be permissive in the sense explained above, it needs to be restrictive in another sense. The nomic account of naturalness cannot get off the ground if lawhood never divides necessary equivalents. Any law *s* is equivalent to countless statements that mention unnatural properties—for example, *s* is equivalent to *s-or-r*, where *r* is some contradiction mentioning *grue*.

So, for the nomic account to work, there need to be permissive fine-grained lawhood facts preceding naturalness facts in the order of fundamentality. But the existing accounts of (permissive) lawhood that don't rely on naturalness do not yield a sufficiently fine-grained notion. Take for instance a simple reductionist view of non-fundamental lawhood (Oppenheim and Putnam 1958): *s* is a law if and only if it follows from the laws of physics together with some definitions of the terms in *s*. This notion of lawhood is too coarse-grained for our purposes: if *s* follows from the laws of physics together with some definitions, so does *s-or-r* (regardless of what predicates feature in *r*). The same goes for accounts of non-fundamental laws in terms of stability under a suitably wide class of counterfactual suppositions. If *s* is counterfactually stable in this sense, so is any logically equivalent statement.

As it is familiar from Goodman's work (1955), this kind of problem cannot be solved by stipulating that only non-disjunctive statements are to be considered. Suppose that ‘All emeralds are green’ and ‘All sapphires are blue’ both follow from the fundamental laws together with definitions of the relevant terms. Then ‘All gremeralds are grue’ will also follow from these laws together with some definitions (where an object is a ‘gremerald’ if and only if it is either observed before 3000 and an emerald, or not observed before 3000 and a sapphire). Our challenge is to find an account of lawhood that rules out statements involving the wrong predicates without invoking the notion of naturalness that we are trying to define.

In his paper about the autonomy of the special sciences, Jerry Fodor favorably discusses the nomic account of naturalness, but alludes to this potential circularity:[12]

> If I knew what a law is, and if I believed that scientific theories consist just of bodies of laws, then I could say that P is a natural kind predicate relative to S iff S contains proper laws of the form $Px \rightarrow Ax$ or $Ax \rightarrow Px$ [. . .] I am inclined to say this even in my present state of ignorance, accepting the consequence that it makes the murky notion of a natural kind viciously dependent on the equally murky notions law and theory. There is no firm footing here. If we disagree about what is a natural kind, we will probably also disagree about what is a law, and for the same reasons. (Fodor 1974)

Because of this circularity, Fodor was tempted by the view that facts about permissive lawhood cannot be reductively explained. This sort of nomic primitivism escapes the circularity worry just discussed, but at too high a cost: like primitivism about naturalness, it leaves little room for systematic metaphysical explanations of the world's macro-structure in terms of micro-physics.

[12] Here I'm glossing over the distinction between natural kinds and natural properties, in part because I suspect that Fodor's use of ‘natural kind’ roughly corresponds to my use of ‘natural property’. In many other contexts, the two notions are importantly different. In particular, traditional accounts of natural kinds primarily target classes of objects to which bundles of natural properties stably attach, in a way that suggests the presence of an underlying common essence (e.g., ‘lion’ or ‘water’).

3.4 | Desiderata

Our discussion so far can be summarized in terms of four desiderata on the notion of lawhood invoked by the nomic account of naturalness:

1. **Permissiveness**: There should be enough laws to cover statements mentioning all the non-fundamental natural properties.
2. **Fine-Grainedness**: The set of laws should not be closed under logical entailment.
3. **Non-Circularity**: The lawhood of statements mentioning a property should not metaphysically depend on that property's being natural.
4. **Derivativeness**: The lawhood of any statement mentioning a non-fundamental property should be metaphysically derivative.

The reason that a primitivist conception of lawhood can't save the nomic account of naturalness is that it cannot satisfy both permissiveness and derivativeness. Existing reductive accounts of permissive lawhood satisfy those two desiderata as well as non-circularity, but they violate fine-grainedness. To make things worse, the project of reducing fine-grained lawhood while respecting non-circularity may seem hopeless. I will now try to convince you that it is not, by sketching a new version of the best system account of lawhood (Mill (1843), Ramsey (1978)[fp. 1928], Lewis (1973)) that meets these four desiderata, and delivers a notion of naturalness that avoids the problems for primitivism.

4 | A BEST SYSTEM ACCOUNT OF PERMISSIVE LAWHOOD

4.1 | Best Systems

David Lewis famously held that natural laws are statements pertaining to the axiomatic system that best summarizes the distribution of fundamental (non-modal) properties throughout space and time. The 'best system' is an accurate (partial) description of the actual distribution of properties, which strikes the optimal balance of competing theoretical virtues, such as simplicity and informativeness (Lewis, 1973).[13]

Two features of the best system account are of special interest, given our target notion of lawhood. First, in line with the derivativeness desideratum, best system laws are not fundamental ingredients of reality. Second, the best system account has a sort of fine-grainedness built into it. As simplicity is one of the selection criteria, only systems consisting of a limited number of axioms make good candidates. If we reserve the term 'law' for the axioms of the best system, we needn't worry about ending up with systems of laws that are closed under logical entailment or logical equivalence.[14]

[13] For the reasons mentioned before, we are not restricted to sets of sentences in our own human languages; we imagine a competition that takes place among sets of sentences in some suitable abstract interpreted language (more on what makes for suitability later).

[14] Lewis uses 'law' to cover all the statements that follow from the axioms of the best system. Nothing important hangs on this terminological choice. Once we have the resources to distinguish between axioms and theorems, we can escape the over-generalization worry by tying naturalness to the axioms.

Traditionally, the best system account has been associated with a T-shirt conception of lawhood, on which the laws of our world are a few equations simple enough to all fit on the front of a T-shirt.[15] Here I'm after a more permissive notion, so some aspects of the account will need to be reworked.

Lewis seems to have thought that a system fitting the T-shirt ideal would be the winner on his view. But note that, until we have specified a particular balancing function (or a set of admissible balancing functions), there is little reason to expect this. Depending on the relative value we place on simplicity and informativeness, more or less expansive systems will count as 'best'.

But how are we to justify a choice of balancing function? Plausibly by reference to the role that laws play in the parts of science that invoke them (fundamental physics, in the case of T-shirt laws). We want our balancing function to yield the kind of best system that is well-suited to play the scientific role of laws. The conjecture that motivates this project is that there is an important theoretical role in non-fundamental science for a notion of 'best system' that is tied to a vastly more permissive balancing function than T-shirt lawhood—perhaps one that allows for enough axioms to fill an entire library.[16]

Since our target notion of lawhood is different from Lewis's, a couple of adjustments to his conception of the virtues are needed. The first has to do with the possibility of laws that are not strictly true. The second—and most significant for our purposes—has to do with the virtue of simplicity. I will consider each one in turn.

On Lewis's version of the best system account, only sets of true statements count as candidate systems. Now, as is widely recognized, the special sciences are rarely in the business of searching for strict, exceptionless generalizations. In light of this, the Lewisian idea that only sets of true axioms make candidate systems seems unmotivated. There is more than one way to accommodate exceptions in the best system account (Braddon-Mitchell (2001), Schrenk (2006)); my preferred strategy is to appeal to a third virtue – 'accuracy' – which can be traded off against simplicity and informativeness. This is perfectly consistent with the general conception of laws as summaries: summarizing effectively often requires making simplifying assumptions that are close enough to the truth but not strictly true.[17]

Apart from tolerating exceptions, many special science theories use loose terms to capture tendencies—e.g.,'Fs are Gs', 'Typical Fs are Gs', '$f(X, Y)$ approximately equals Z'.[18] While standard best system accounts do not cover such generalizations, I see no reason to think they couldn't be accounted for in this framework. Such statements might lack precise truth-conditions, but their

[15] This 'T-shirt constraint' is credited to the physicist Leon Max Lederman (1993): "My ambition is to live to see all of physics reduced to a formula so elegant and simple that it will fit easily on the front of a T-Shirt."

[16] This does not require giving up on the T-shirt conception of lawhood: we may need one kind of best system account for an important conception of lawhood that fits the T-shirt ideal, and also a modified best system account of another important notion of lawhood that plays an analogous role in the special sciences.

[17] For some preliminary suggestions on how to define the needed notion of accuracy, see Gómez Sánchez (2020). For simplicity, I will assume that the accuracy of a system is evaluated across spacetime, but it may be more appropriate to think of the special sciences as seeking to summarize local spatio-temporal regions that are of particular interest to us (Mitchell 2000). To account for this in the best systems framework, we can relativize accuracy to spatio-temporal regions, yielding a region-relative notion of lawhood (see Gómez Sánchez, 2020).

[18] Consider, for example, the statement that enclosed gases approximate the ideal gas equation, or the statement that humans are approximately rational in typical exchanges of goods. These generalizations play the role of laws in many scientific explanations, and it is doubtful that those explanations could be recast as strict generalizations without loss in simplicity or informativeness.

meanings may determine how accurate they are at various worlds. If so, then these statements could enter the best system competition.[19]

There is a lot more to say about accuracy and its relation to exceptions, idealization, and non-strict laws. My aim has been only to explain why exceptions and idealizations are not in-principle barriers to the project of analyzing permissive lawhood in terms of best systems. From now on, I will abstract away from this issue, in order to focus on an orthogonal issue which is much more pressing in the present context: whether there is a tenable conception of the virtue of simplicity that does not invoke naturalness.

4.2 | Simplicity and permissive lawhood

For Lewis, only sets of sentences stated in a fundamental language are candidate systems. Systems mentioning properties like *spinach*, *person*, or *city* cannot even enter the competition.[20] In giving a best system account of permissive laws, we need to somehow broaden the class of candidate systems. Yet, as the following two arguments will show, the best system account is hopeless unless some vocabularies are privileged over others in the best system competition. After presenting these arguments, I take on the challenge of stating the needed restriction on vocabulary without invoking naturalness.

The first argument is well-known, because it played an important role in Lewis's own thinking about simplicity and naturalness. Lewis considers an atomic predicate 'F' which applies to everything in the actual world, and to nothing in other worlds. This predicate figures in a single-axiom system that threatens to better any of the systems that we would regard as serious contenders. Consider some such contender, and name it S. S cannot be any more accurate than $\forall xFx$ relative to the actual world, because $\forall xFx$ is already maximally accurate. Secondly, $\forall xFx$ seems to be no less informative than S.[21] And thirdly, S is surely more complex than $\forall xFx$ on any syntactic conception of simplicity. So, unless this system gets penalized for employing the funny predicate 'F', $\forall xFx$ will be the only law, and F the only natural property.

This argument shows that we need some kind of restriction on admissible predicates. The problem for us is to find a restriction on predicates that, unlike Lewis's restriction, rules in non-fundamental predicates but can be stated without appeal to naturalness.

Lewis's argument is decisive, but focusing on it alone might mislead us into thinking that a minimal restriction would suffice—e.g, restrictions to qualitative predicates, predicates that always divide world-mates, or ones that have instances in more than one world. For this reason, I would like to offer a more general argument, which is a variant of Putnam's permutation argument for

[19] In fact, Lewis's strategy for dealing with probabilistic laws can serve as a model for a best system account of non-strict laws (Lewis, 1994). Here is a sketch: we start by expanding the vocabulary that best systems can draw from, to allow for candidate axioms such as: 'Generic Fs are Gs' and/or 'Almost all Fs are Gs'. The loose terms 'generic' and 'almost all' are uninterpreted at this stage—so the corresponding statements don't have truth-conditions. We then provide general rules for assessing the accuracy of statements involving each of our new terms—e.g., 'generic', 'almost all'. For example, we might stipulate that 'Almost all Fs are Gs' is highly accurate when all but one Fs are Gs, a little less accurate when all but 2 Fs are Gs, and so on.

[20] There may be sentences in FL that have the same truth-conditions as 'Spinach is iron-rich', but having such statements in the best system wouldn't guarantee the naturalness of *spinach* by the lights of my account.

[21] To properly defend this step in the argument we need to make some assumption about informativeness. Here's one assumption that would do: If the truth of S necessitates the truth of S', but not vice versa, then S' is no more informative than S.

semantic indeterminacy (See Putnam (1981), Chapter 2 & Appendix). In presenting this argument, I will assume world-bound individuals, and a simple-minded absolutist conception of physical quantities as relations between physical objects and real numbers; the argument strategy does not require either of these assumptions.

Suppose, for concreteness, that the only actual law of our world is $F = ma$. Let S be a system whose only axiom is $F = ma$. My Putnam-style argument will show that, given a syntactic conception of simplicity and no restriction on the predicates that feature in candidate systems, there is a system that is no less accurate, no less informative, and no more complex than S—but involves unnatural predicates.

First, we qualitatively define a set of Newtonian worlds where $F = ma$ holds non-vacuously. For example, we might stipulate that W stands for the set of Newtonian worlds containing only two particles. Next, we define *schmass* (or m^*) as follows: a (world-bound) possible individual a has schmass x if and only if a doesn't live anywhere in W and its mass is x or a lives somewhere in W and its mass is x^2. *mass* and *schmass* have intensions that coincide outside of W and diverge everywhere in W.

Now let S^* name a system whose only axiom is $F = m^*a$ (rather than $F = ma$). Given the way *schmass* was defined, the system S^* will have the same truth value as S everywhere outside of W. Now, since every world in W is (non-vacuously) Newtonian, there are some objects in every world in W whose force, mass and acceleration conform to $F = ma$, and hence violate $F = m^*a$ (when F is non-zero). Thus, $F = m^*a$, unlike $F = ma$, is false inside W (assuming non-zero forces act on the particles in W).

The problem is that S^* is no worse than S with respect to accuracy, informativeness or syntactic simplicity. $F = m^*a$ must be just as accurate as $F = ma$, since *mass* and *schmass* don't differ in extension. Since $F = ma$ and $F = m^*a$ are syntactically alike, only informativeness remains to break the tie. Yet it is doubtful that S will have an edge over S^* in this respect since, by construction, $F = m^*a$ rules out strictly more possibilities than $F = ma$. Moreover, a pair of scientists using S and S^* respectively would end up with equally accurate predictions about actual (and also nearby) worlds. Unless we were to somehow bring naturalness into the competition (thus violating the non-circularity desideratum), it is hard to see how S could outcompete S^*. Thus, liberalizing the best system competition to allow for systems stated in any languages yields the problematic prediction that S^* is no worse than S, failing to accommodate the obvious fact that *mass*, not *schmass*, is natural.

This second argument shows in another way what Lewis had shown already: the best system account is hopeless if simplicity is merely syntactic and there is no way of privileging some vocabularies over others. But note that *schmass* cannot be ruled out as easily as Lewis's trivial property F: *schmass* is qualitative, covers objects across many worlds, and often divides worldmates.

Like Lewis, I think that appealing to the distinction between fundamental and non-fundamental properties is the only promising way forward. But rather than imposing a blanket ban on all non-fundamental predicates, I want to explore another conception of how fundamental structure constrains simplicity assessments. Systems with problematic predicates like F or 'schmass', I will suggest, fail to be simple in the sense that matters for lawhood: their syntactic simplicity hides an underlying 'semantic complexity', which is revealed when we unpack their meaning in more fundamental terms. The next section shows how to make this idea precise without violating the non-circularity desideratum.

4.3 | Semantic Complexity and Multiple Realizability

Suppose we were in a Newtonian world whose fundamental properties and relations include *mass* but not *schmass*. Then there would be an obvious asymmetry between $F = ma$ and the wacky systems considered above. The wacky systems mascarade as simple by using predicates whose definitions in more fundamental terms are rather complex. If these systems had to pay a complexity penalty in proportion to the FL-complexity of their predicates, we would expect the kinds of systems that physicists take seriously to come out far ahead.

In a different context, Hicks and Schaffer (2017) consider an amendment to Lewis's best system account along these lines. To allow for physical laws that mention non-fundamental quantities like derivatives and sums of forces, Schaffer and Hicks consider a view on which a system is assessed not only with regards to its syntactic simplicity, but the FL-complexity of the predicates of that system.[22] We can think of this proposal as recognizing two dimensions of the virtue of simplicity: a syntactic dimension and a semantic dimension.

This amendment to the account is a step in the right direction, but does not yet deliver the permissive notion of lawhood that I'm after. The possibility of good systems involving multiply realizable properties would remain mysterious, unless we had good reason to expect the FL-definitions of such properties to be relatively simple.

In discussing Lewis's account of naturalness in terms of FL-complexity, I had remained open to the possibility that multiply realizable properties have relatively simple functional definitions, that is, definitions that appeal to those properties' causal or nomic roles. For example, we might define *pain* along these lines: to be in pain is to have some property or other that figures in laws relating it to inputs, other mental properties, and behaviors in such and such ways. (This would have to be filled out by a detailed description of the kinds of laws that characterize 'the pain role', and a specification of the place of pain-realizers in those laws.)

But note that, even granting that all multiply realizable properties are so definable, it does not follow that they have simple enough FL-definitions. Firstly, even if a fundamental language contains nomic vocabulary, this nomic vocabulary is not going to include a term for permissive lawhood, a non-fudamental notion. Secondly, the laws about realizers will not be easily describable in FL if they mention non-fundamental properties that lack simple FL-definitions.

What we need is a notion of semantic simplicity on which properties with simple functional definitions in terms of 'lower-level' laws get counted as semantically simple, even if those laws are not laws in a fundamental sense of lawhood, and they invoke non-fundamental terms. The rest of this section develops a proposal along these lines that gets around the apparent circularity of invoking lawhood in an account of lawhood.

Rather than running one best system competition, imagine running many iterations of it, adjusting the winning criteria slightly at each stage. In the first stage, we follow Schaffer and Hicks' recipe: we reward systems with lower 'semantic simplicity' which is measured by adding the FL-complexities of the primitive predicates in a system. This yields a first system of laws, which doesn't yet mention any multiply realizable properties. We may call the axioms of the system that wins this first competition 'basic' or 'level$_1$' laws.

A more permissive system of laws comes into sight if we consider re-running the best system competition, having slightly adjusted the winning criteria. We now tie semantic complexity to a new 'reference language' RL_1, which results from extending FL by adding simple predicates for

[22] In fact, the view they consider invokes degrees of naturalness rather than FL-complexity directly, but—at least in this paper—they seem to be open to the Lewisian idea that low FL-complexity makes for naturalness.

all the properties mentioned by basic laws, and a lawhood predicate that picks out the property of being a basic law.[23]

The semantic complexity of a system at this second stage will equal the sum of the complexities of the RL_1-definitions (rather than FL-definitions) of the properties in the system. Assuming that RL_1 can quantify over properties and interpreted sentences (or structured propositions), it will have the resources to state functional definitions, allowing systems mentioning functional properties to count as simple.

Continuing this process, our reference language gets expanded at every stage to include predicates for all the properties that feature in laws at lower levels, and a predicate law$_n$ that covers lower-level laws. After carrying out this process infinitely many times, we end up with a stratified system of laws, where each axiom can be assigned a level corresponding to where it first appeared. A statement is a law if and only if it appears somewhere in this stratified system.

The full iterative best system account, stated more precisely, invokes three key notions defined below:

> **Candidacy**: S is a candidate system relative to a reference language RL if and only if S is a set of sentences in a language RL^+ that results from extending RL with finitely many new predicates that have RL-definitions.
>
> **Semantic Complexity**: If S is a candidate system relative to RL, its semantic complexity relative to RL is the sum of the complexities of the (simplest) qualitative RL-definitions of the predicates in S.
>
> **Better System**: S is a better system than S' relative to RL and world w if and only if S achieves a better balance than S' of (i) accuracy relative to w, (ii) informativeness, and (iii) simplicity relative to RL, where the latter is a quantity that monotonically decreases with both syntactic and semantic complexity (relative to RL).

In terms of these three notions we can characterize a class of recursive functions that 'build' the full hierarchy of systems for a world:

> **Lawbook-Constructing Function**
>
> A function f from natural numbers and worlds to sets of interpreted sentences is 'lawbook-constructing' if and only if:
>
> There is a fundamental language FL such that:[24]
>
> (i) For every world w, $f(1, w)$ is the best candidate system with respect to FL and w,
>
> (ii) For every $n \geq 1$ and world w, $f(n+1, w)$ is a candidate system relative to a language RL_n that is better than any other candidate relative to RL_n, where RL_n is a language that results from extending FL by addition of all the predicates in $f(1, w), \ldots, f(n, w)$ as well as a predicate 'law $_n$' whose extension, at each world v, contains all and only the axioms in $f(1, v), \ldots, f(n, v)$.

We finally define lawhood in terms of lawbook-constructing functions as follows:

[23] Recall, from before, that I think of laws as abstract set-theoretic entities coupled with interpretations. So the lawhood predicate would be a two-place predicate meaning: *s* is a first-level law under interpretation *I*.

[24] I'm assuming here that the same notions are fundamental in all worlds. A few adjustments to this definition would be needed to accommodate contingency in the fundamental structure.

Permissive Lawhood: s is a law at w if and only if, for some lawbook-constructing function f and some n, s is a member of $f(n, w)$.

With the full account on the table, let me illustrate, using a toy functional definition of a high-level property, how this account makes room for multiply realizable natural properties.

Let us imagine that the property *hunger* is realized in humans by the activity of AgRP neurons, and in martians by the activity of AgRQ meurons. And suppose that, at the n-th stage of the best system construction, we have the following laws about hunger-realizers in humans and martians respectively:

ψ_{human}: If x is human, and the average firing rate of x's AgRP neurons is r at t, then x's short-term energy stores are below $f(r)$ at t-ε.

χ_{human}: If x is human, and the average firing rate of x's AgRP neurons is r at t, then the strength of x's desire to eat at $t + \varepsilon$ is $g(r)$.

$\psi_{martian}$: If x is a martian, and the avereage firing rate of x's AgRQ meurons is r at t, then x's short-term energy stores are below $f(r)$ at t-ε.

$\chi_{martian}$: If x is a martian, and the average firing rate of x's AgRQ meurons is r at t, then the strength of x's desire to eat at $t + \varepsilon$ is $g(r)$.

These axioms mention other multiply realizable properties, such as energy-stores, and desires. For now, I will assume that these functional notions are already in our reference language RL_n—i.e., that they have already been defined in terms of their role in lower-level laws and earned a place in some laws. As we will later see, the account can also handle cases where clusters of functional properties get defined at once, by reference to the same low-level laws.

At stage $n + 1$ of the best system construction, we assess semantic complexity in a reference language RL_n that extends FL with all the predicates that have appeared in laws up to level n, as well as a lawhood predicate law$_n$ that covers all the laws up to this level. In this reference language, we may define a functional notion of *hunger* along these lines:

hungry $=_{def}$ x instantiates some property Y that figures in ψ-type and χ-type laws$_n$, in such and such syntactic positions.

This definition would have to be further spelled out by describing the two common forms that laws$_n$ connecting hunger-realizers to energy stores and desires take. This is plausibly achievable in RL_n. By construction, RL_n includes the required predicate law$_n$. We just need the further assumptions that, in RL_n, we can quantify over properties, structured propositions (or their sentence-like surrogates), and also that we can easily describe involvement or aboutness relations between the two. (Since RL_n is built upon FL, this imposes certain constraints on FL, but these constraints don't strike me as particularly problematic.)

If *hunger* has this sort of definition in RL_n, it can figure in candidate systems that compete in iteration $n + 1$ at a relatively low simplicity cost. At this stage, we may see generalizations such as 'Hunger decreases motivation for strenuous activity', 'Hungry predators take higher risks while hunting' in the best system.

Note that the definitions of functional properties that count towards semantic complexity will always mention a restricted notion of lawhood: one that covers laws up to a particular level. This allows for the possibility that multiple hunger-like properties with similar nomic roles but tied to different levels appear in the best system at different stages. This is a little odd, but I don't see it as a problem. Our concept of hunger may be indeterminate between many of these (roughly coincident) notions, or it may express the property of having one of these level-specific hunger-like properties.

It is also worth noting that, on this approach, functional notions do not figure in laws parallel to the ones mentioned in their nomic roles. In the above example, we might have expected to find a law saying that, if x is human and hungry to degree d at t, then x's short-term energy stores are below $h(d)$ at t-ε. But, assuming my toy account of *hunger*, this statement is not going to be informative enough to make it to the best system, for we can work out that it holds (or that it is highly accurate) just by knowing the definition of *hunger*. Relatedly, the account rules out natural properties that figure in no interesting generalizations beyond those mentioned by their functional definitions. As Fodor (1983) put it, a natural property corresponds to 'a class of phenomena that have many scientifically interesting properties in common over and above whatever properties define the class'.

I earlier assumed that all the properties mentioned by my toy definition of *hunger* (e.g., desire and short-term energy stores) had already appeared in laws at levels n or below. This required that they have definitions that don't mention *hunger*. A corresponding assumption may seem problematic in other cases. For example, perhaps any plausible functional definition of belief will have to mention desire-like states, and any plausible functional definition of desire will have to mention belief-like states. While my account does not allow for mutual reference in functional definitions, it accommodates something close enough. Let me illustrate with the case of belief and desire.

Suppose that, at levels n or below, there are a series of kind-specific laws, each one connecting desire-realizers, belief-realizers and behaviors (in the same way): $\phi_{humans}, \phi_{dogs}, \phi_{martians} \ldots$ Corresponding to these laws would be a joint 'belief-desire nomic role'. Two relations X, Y play this nomic role for kind K if and only there is a ϕ-type law$_n$ that features predicates for K, X and Y in the appropriate syntactic positions.

In terms of this role, we can give Lewis-style functional definitions of belief and desire as follows:

> x believes(/desires) that $p =_{def}$ for some kind K and relations X and Y, x is a K, x bears relation $X(/Y)$ to p, and X and Y play the belief-desire role in some laws$_n$ for kind K.

Using this method, it is in principle possible to jointly define all mental predicates at once. However, this makes the definition of every mental predicate very complex, which—given the present account—makes it less likely that they will end up in the best system. Thus, my account more plausibly delivers the naturalness of mental predicates given a version of functionalism on which each mental predicate is defined independently of most others, with some local constitutive connections.

While the paradigms of multiple realizability are mental properties, many non-mental properties are also better accounted for within the iterative best system account. To use a familiar example, suppose that 'All emeralds are green' and 'All sapphires are blue' are, from the perspective of accuracy and informativeness, good candidates to feature in a permissive best system.[25] The *prima facie* challenge for a best system account of permissive fine-grained lawhood is that

[25] To a first approximation, it seems likely that similar generalizations involving more specific color properties (say, *emerald-green* and *sapphire-blue*) would make for better axioms by virtue of being more informative. But, given that the best system account is holistic, determining which way of carving color space is most natural in our world would require

there are gruesome correlates of these statements that are similarly accurate and encode the same information: 'All gremeralds are grue' and 'All grapphires are bleen'.

Since syntactic simplicity alone does not favor the right systems over the gruesome ones, we need the notion of semantic complexity to ground the desired asymmetry. Now, if we were evaluating the semantic complexity of these statements by reference to FL-definitions, we would struggle to explain how any color properties can be natural. There is no simple micro-physical property that is necessary and sufficient for being a certain color; to see what is common between all the green things, for example, we must take into account their functional profiles (in particular, their dispositions to interact with light in certain ways). But, for similar reasons to those we encountered with *pain* or *hunger*, the relevant functional profile seems unlikely to have a particularly simple description in fundamental terms.

My iterative account is better suited to explain the green/grue contrast. For green might have a functional definition roughly along these lines: x is green if and only if x has some property Y that features in a law$_n$ of the form 'Ys typically reflect light with such and such wave-properties under such and such lighting conditions'.[26] If some such definition is adequate, then statements like 'All emeralds are green' (and other color generalizations) could feature in a high-level best system. And, while roughly the same information could be encoded in gruesome vocabulary, competing systems with gruesome terms will lose out because they use predicates with more complex definitions in our reference language; for instance, the simplest definitions of grue in terms of lower-level laws and lower-level natural properties would arguably have the functional definitions of *green* and *blue* as proper parts.

Let us review where we've got to. We had previously arrived at four desiderata for our notion of lawhood: (i) that it be permissive enough to cover non-fundamental properties, (ii) that it be fine-grained, (iii) that it not make the lawhood of a statement dependent on the naturalness of its predicates, and (iv) that it treat lawhood facts involving non-fundamental properties as derivative. We had seen already that any best system account is well-placed to satisfy desiderata (ii) and (iv). My version of the account is designed to reconcile the remaining two desiderata. It does this by letting non-fundamental predicates figure in candidate systems, when they are easily definable in terms of lower-level properties and laws. We exclude gruesome alternative systems by imposing the constraint that our hierarchy of systems bottom out in a fundamental language in a certain way, where the notion of a fundamental language gets characterized in terms of fundamentality (rather than naturalness).

Unlike primitivism, the conception of laws and naturalness developed here is consistent with PURITY and ATOMISM, because it does not take the naturalness of non-fundamental properties as brute. Moreover, it recovers the explanatory connections across levels of description that primitivism fails to account for. Recall the example of the naturalness of the property *H_2O molecule*, which does not seem independent of the way it is grounded in more fundamental properties (e.g., *hydrogen*, *oxygen*, *bond*…). The present account accommodates this. To metaphysically explain how *H_2O molecule* ends up being natural, we must explain how it ends up in the best system. And the fact that it figures in this system itself depends on the fact that it has a relatively simple definition in more fundamental terms which figure in laws at previous levels of the hierarchy (if *oxygen* and *hydrogen* had not been natural, then *H_2O molecule* wouldn't have been natural either).

looking at many different generalizations involving color terms (including generalizations in perceptual psychology). I abstract away from such complications in what follows.

[26] See Cohen (2003) for a more sophisticated functionalist account of color.

This iterative best system account of lawhood serves as an existence proof for the joint satisfiability of the desiderata I had placed on a notion of lawhood. But the overall strategy I have described is independent of some of the core commitments that originally motivated the best system account. For example, a core commitment of Lewis's view is that the facts that the best system summarizes are all non-modal and non-nomic. But the idea that fine-grained laws are statements that balance various theoretical virtues can be applied in other settings.

Heather Demarest (2017) proposes that laws are statements that summarize all the truths that flow from the essences of the fundamental properties, a primitivist about nomological necessity can consider laws to be summaries of the nomological necessities, and a primitivist about counterfactuals could consider laws to be summaries of patterns of counterfactual dependence. In a similar spirit, I have suggested in other work that we think of (permissive) laws as summarizing sets of physically nearby worlds ('modal neighborhoods'). My proposed way of tying the virtue of simplicity to fundamentality via an iterative best systems construction will be available on any of these approaches.

5 | THE ROLE OF NATURALNESS

In the rest of the paper I want to argue that the notions of naturalness and fine-grained lawhood defined in the previous sections are well-suited to play the three key roles that naturalness has been invoked to play: in the theory of causation/explanation, in the theory of rational induction, and in the theory of meaning/reference. I will then go over some other roles that are best played by the closely related notion of fundamentality.

5.1 | Causation/Explanation

Right now, thousands of traffic-lights are about to transition from red to green, causing the cars in front of them to accelerate soon after. Each of those traffic-lights is also about to undergo another transition: from red to grue. But this other transition does not seem to enjoy the same causal status as the former one: causally explaining the movement of a car in terms of the grue light in front of it would miss the situation's objective structure.[27]

In order to capture this intuitive judgment, standard theories of causation/causal explanation appeal to the notion of naturalness. Take the simple counterfactual account of causation: C causes E if and only if C and E are actual distinct events, and E wouldn't have occurred if C hadn't. Let C_1 be the event of the traffic-light going green at t, and let C_2 be the event of the traffic-light going grue at t. Note that $\neg C_1$ and $\neg C_2$ are 'subjunctively equivalent'—that is, $\neg C_1 > \neg C_2$ and $\neg C_2 > \neg C_1$ are both true. (The closest worlds where the traffic-light does not go green at t are worlds where it doesn't go grue at t, and vice versa). It follows from this, given standard counterfactual semantics, that:$(\neg C_1 > \neg E) \leftrightarrow (\neg C_2 > \neg E)$. So E counterfactually depends on the traffic-light going green if

[27] I'm inclined to disregard the distinction between causation and causal explanation in this context, as I don't think it is metaphysically significant. But those who care about this distinction may want to allow for the event involving *grue* to be a cause of E, without allowing *grue* to feature in E's causal explanation. Readers who are sympathetic to that approach should read the present sub-section as drawing a connection between naturalness and causal explanation (rather than causation).

and only if it also counterfactually depends on its going grue.[28] To avoid the result that C_2 has the same causal status as C_1, Lewis denies that C_2 is an event at all, by analytically tying the notion of an 'event' to natural properties.[29] The same dialectic replays itself in the context of interventionist theories of causation. As Laura Franklin-Hall (2016) has persuasively argued, interventionist accounts struggle to make good on their promises unless they appeal to a notion of naturalness.

Many have found such an appeal to naturalness objectionable. A possible source of concern is the impression that the notion means nothing more than 'causally potent' (or: 'capable of featuring in causal explanation'), which would put into question the reductionist credentials of any theory of causation that appeals to it. By adopting the nomic account of naturalness we can break out of the circle: on this view, we define naturalness in terms of lawhood (not causation, or 'causal potency'). From this perspective, there's nothing particularly problematic or mysterious about the connection between causation and naturalness that Lewis draws. Causal explanations trace counterfactual patterns, as seen through the lens of a nomically privileged conceptual repertoire.

We have seen one concrete proposal for vindicating the intuition that gruesome properties are absent from causal explanations. But the nomic account of naturalness developed here also fits well with another proposal. If causation reduces to fine-grained lawhood, we may be able to give a deeper explanation of why unnatural properties fail to be causally relevant.

Consider a simple covering-law account of causation/causal explanation.[30] An object a's being F at t (Fa_t) causally explains an object's b being G at $t + \varepsilon$ ($Gb_{t+\varepsilon}$) if, for some background condition B and relational predicate R,

(i) '$\forall x \forall y \forall t \& ((Fx_t \& Rxy_t \& B_t) \rightarrow Gy_{t+\varepsilon})$' is a causal law, and
(ii) $Fa_t \& Rab_t \& B_t$.

Assuming a fine-grained conception of laws, traffic-lights turning grue would not figure in any law, and similarly for other putative causal explanations invoking unnatural properties. Thus, this kind of account would preclude unnatural properties from featuring in causal explanations.

Covering law views have fallen out of favor for a number of reasons. Some of these reasons are less decisive if we have a permissive notion of lawhood. For example, one worry is that properties like 'being a traffic-light' or 'being a car' don't feature in any laws at all, meaning that the explanation for cars accelerating in terms of green lights won't be captured by this kind of account (Paul and Hall 2013). But, on the present view, there may well be lawful regularities about cars and traffic-lights in the permissive sense of that notion.

Other challenges for the covering law view remain. For this account to work we need to be able to non-circularly define 'causal law', and we need this notion to be narrow enough so as to exclude, for example, robust generalizations that run in the wrong direction, such as generalizations saying that if a certain symptom is present then there is (probably) a certain disease underlying it. Despite these challenges, some of us still see potential in the general idea that all causal relations are

[28] By the principle: $\phi_1 > \phi_2, \phi_2 > \phi_1, \phi_1 > \psi \vDash \phi_2 > \psi$. This principle is validated by both Lewis and Stalnaker's counterfactual semantics.

[29] I've been speaking of events 'involving' properties. I should say that, for Lewis, events are just sets of possible spatiotemporal regions, so they don't have any properties as constituents. However, we can make sense of a notion of property involvement nonetheless, since the sharing of (natural) properties plays a role in grounding the fact that a set of regions is an event.

[30] See Davidson (1967) for an early version of the covering-law view. For more recent versions of this kind of view, see Cartwright (1999), Maudlin (2004), and Strevens (2008a).

underwritten by fine-grained laws of a certain sort. And if that idea pans out, we might get a nice explanation for why gruesome properties fail to feature in causal explanations: they don't figure in laws.

The nomic account also has the potential to illuminate the connection between naturalness and explanation in non-causal domains. For example, we may consider a best system account of mathematical 'laws', which would give us a notion of naturalness for mathematical notions. If we think of mathematical explanations as special kinds of derivations of mathematical facts from the axioms, then we can glean an explanation for why gruesome mathematical notions (e.g., quaddition) feature in no explanations.[31] Even the role of naturalness in metaphysical theorizing may be illuminated by the nomic account. If metaphysicians are in the business of constructing a best system of a certain kind, and if the generalizations that figure in this system back explanations, we should expect natural notions to be privileged by these explanations too.

5.2 | Naturalness and meta-semantics

The second central role for naturalness is in the theory of meaning. Lewis (1983) appealed to naturalness to avoid indeterminacy worries with interpretivist theories of meaning. In what follows, I argue that my account of naturalness helps demystify the connection to reference that Lewis posited, and I explain how other theories of meaning may be able to draw on the tools provided in this paper to solve parallel indeterminacy issues.

Interpretivism is the view that the meanings of someone's words and mental representations are given by the interpretation function that a charitable interpreter would arrive at, if she had access to all of the speaker/thinker's linguistic and non-linguistic behavior. To count as 'charitable', an interpreter must assign truth-conditions to the person's utterances/ thoughts in a way that maximizes their internal coherence and/or truth (Davidson (1973), Lewis (1974)).

But charity (so conceived) fails to pin down determinate meanings for a subject's language. Let *f* be the interpretation function that assigns the correct intensions to predicates and the correct referents to names in a given language—i.e., the intensions and referents that give the actual meanings for the expressions of that language. As Putnam (1981) observed, *f* is not uniquely recommended by charity reasoning. Consider a pair of names in the relevant language, such as 'Obama' and '1'. Now construct a function *f** which is almost exactly like *f*, but 'permutes' these two objects systematically. *f** assigns the number 1 as referent to the name 'Obama', and Obama to the name '1'. It includes the number 1 (instead of Obama) in the intension of 'person', and Obama in the intension of 'integer'. Since *f** permutes these two individuals systematically, the truth-values of all the sentences in the language will be preserved under f^*. For example, 'Obama is a person' is true because the referent of Obama under f^* (i.e., the number 1) falls under 'person' as interpreted under f^*. Hence, f^* is guaranteed to be just as charitable as f: the same beliefs and utterances come out true on both interpretations, and whatever is consistent under f will also be consistent under f^*.

Lewis (1984) suggested that the permuted interpretations could be ruled out by appeal to naturalness. The correct interpretation for a given language, according to Lewis, is the one that best balances charity and naturalness. Because the permuted interpretation f^* maps predicates

[31] Quaddition (Kripke 1982) is a mathematical function that resembles addition for inputs up to 57, and then diverges.

like 'person' and 'number' to highly unnatural properties, it will do poorly by the lights of the naturalness desideratum.[32]

Because Lewis tied naturalness to *FL*-complexity, he had trouble explaining why multiply realizable properties like *pain*, which lack simple *FL*-definitions, are highly eligible meanings for our terms. This problem is particularly pressing in the light of my observation in §1.2 that there are ways of contracting the *FL*-definition of *pain* which correspond to less natural properties. Since some of these properties would have a similar extension to *pain*, they would likely be reasonably good at capturing use. If so, Lewis would predict that our term 'pain' would express one of these properties.

Natural properties, as conceived under the nomic account, make suitable reference magnets. This view can explain why certain multiply realizable properties are highly eligible meanings despite lacking simple *FL*-definitions, because they are privileged by the nomic structure of the world. And, although it is doubtful that all of our simple predicates express properties that figure in laws, it is plausible that a property's eligibility has to do with the simplicity of its definition in natural terms.

Moreover, my version of the nomic account nicely complements a proposed amendment to interpretivism due to Robert Williams (2007), which promises to demystify the connection between meaning/reference and naturalness. In broad strokes, Williams suggests that an adequate interpretation for a language is the 'best system' of the patterns of use of that language. Reference magnetism would be explained by the fact that simplicity in natural terms is a good-making feature of a summary. Note that this proposal requires a notion of naturalness which applies to higher-level properties (including multiply realizable ones). My iterative account provides the required notion: once a property like *pain* figures in some laws, it becomes highly eligible to figure in other parts of the best system, including statements that summarize use.

The indeterminacy worry that Putnam's permutation argument makes vivid afflicts not only interpretivism, but any account of mental content. Fortunately, my notions of naturalness and fine-grained lawhood could do work other accounts as well, and Williams' proposal for demystifying reference magnetism has parallels in these other contexts.

For concreteness, let us focus on a simple version of informational semantics (based on Dretske's (1981) account). To keep it simple, I will restrict the account to representations of the form F · i, which attribute some property denoted by F to an object denoted by a perceptual index i. The account states, roughly, that the perceptual representation F picks out the most specific property that it tracks during its 'learning period', where a property is tracked by F if it has sufficiently high objective probability of being present in an object, conditional on the subject perceptually subsuming that object under F (by suitably binding F with an index denoting the object).

This account succumbs to Putnam's objection if all the abundant properties are among the candidate meanings for a perceptual representation like GREEN. To see this, consider a slight variant of Goodman's 'grue': to be grue* is to be either green and perceived by some sentient being at some point in time, or blue and never perceived at any time. Given this definition, necessarily, whenever someone perceives a green object, that object is also grue*. It follows that the probability that a visually tracked object is grue* conditional on its being perceptually subsumed under GREEN is just as high as the probability that it is green on the same condition. Therefore, if the representation GREEN tracks *green* during the learning period, it also tracks *grue** (equally well) during that same period. But neither property is more specific than the other, so Dretske cannot explain

[32] Weatherson (2013) argues that Lewis's official view of meaning does not treat naturalness as a separate desideratum: rather, naturalness constrains the rationality of beliefs, which in turn constrains what counts as a charitable interpretation.

why GREEN picks out one rather than the other without some sort of naturalness constraint on candidate meanings.

To avoid this problem, Dretske can borrow Lewis's strategy. He could say that *grue** was never a candidate meaning for GREEN, because only natural properties are eligible meanings for (logically simple) mental representations. Alternatively, he could connect reference directly to fine-grained lawhood. He could require, for example, that there be certain kinds of fine-grained laws connecting mental representations directly to the properties they express. This latter option sits well with a Williams-style defense of reference magnetism: the connection between reference and naturalness would be explained via their respective connections to laws.

I don't think that appealing to naturalness and/or fine-grained lawhood would be enough to save Dretske's informational account of content.[33] Rather than defending this particular account of content, my aim here was to highlight how the appeal to naturalness and fine-grained lawhood can do work in a theory of meaning, and also to remind skeptics of reference magnetism that there need not be any deep mystery in the connection between reference and naturalness —at least not if we grant that reference/meaning is based on fine-grained nomic notions (causation, explanation, probability, or lawhood) and endorse a nomic account of naturalness.

5.3 | Induction

'All emeralds are green' probably strikes you as a better hypothesis than 'All emeralds are grue'. You take your observations of green emeralds to provide inductive support for the former and not the latter. Yet, as Goodman (1955) pointed out, there is a sense in which your observations of green emeralds do not favor one hypothesis over the other: both hypotheses imply that all the emeralds you've seen have been green (given that it is before the year 3000). Moreover, simplicity alone cannot break the tie, for the two hypotheses have the same form.[34]

Goodman (1955) concluded from this case that what we should believe depends not only on what evidence we have, but also on which predicates are 'entrenched' for us—which ones we happen to have used successfully thus far. This leads to the strange conclusion that, upon encountering culturally and/or psychologically different beings who expect to see blue emeralds after the year 3000, we should be willing to epistemically praise their inferences without feeling any pressure to revise our own.

Many of us find this subjectivism hard to accept. Believers in natural properties, in particular, want to point to differences in naturalness to explain the epistemic superiority of some inductive inferences over others (Sider, 2011). The hope is that there is some general epistemic principle connecting induction and naturalness which, when applied to Goodman's case, tips the balance in favor of the green hypothesis.[35]

In a Bayesian framework, questions of inductive support reduce to questions about which priors are reasonable. We might hope that, by appeal to the notion of naturalness, we will be able to formulate a plausible constraint on priors which favors hypotheses that are simple to state in natural terms. But we cannot simply require that, given any pair of mutually exclusive propositions,

[33] He would still have the problem of saying what counts as the 'learning period', and of precisely specifying a probabilistic constraint on reference that is not implausibly demanding.

[34] For a more rigorous and more general presentation of this challenge, see Titelbaum (2010).

[35] Whether this norm would be one concerning rationality, justification or some other notion of reasonableness is a subtle question that I can't do justice here.

the one which is simpler to state in natural terms get higher probability. (We would have, for example, that 'All emeralds are red' is more probable than 'All emeralds are green-or-blue', and also that 'All emeralds are green' is more probable than 'All emeralds are blue-or-red').[36]

Interestingly, a constraint on priors that does justice to our intuitions and avoids these kinds of problems naturally falls out of the nomic account of naturalness developed here, when combined with the popular thought that we are a *priori* justified in expecting that the laws of nature are simple. As I will explain, this constraint is best thought of as defining an ideal of inductive reasoning, which we can't realistically hope to attain. But understanding how induction works in this ideal case is illuminating nonetheless, and an important first step toward understanding rational induction in our own case.

Consider an idealized agent who is logically omniscient, knows a *priori* which properties are fundamental, and can work out a *priori* the real definition of any concept, including lawhood and naturalness. I want to suggest that this agent would do induction by formulating all the possible systems of laws, generating a prior probability distribution over them that favors simpler systems over more complex ones, and then conditionalizing this prior on the totality of her evidence. In what follows I will explain this method more precisely, and argue that it would lead this agent to respond to evidence in ways that strike us as reasonable—for example, that she would learn that all emeralds are green upon observing enough green emeralds. I will then discuss what bearing this observation has for the epistemology of induction more broadly.

Let a 'complete nomic hypothesis' be a proposition that fully describes a world's nomic structure. Intuitively, complete nomic hypotheses are propositions that always divide worlds that differ in their laws, and always unite worlds that have exactly the same laws (and therefore exactly the same natural properties).

Many complete nomic hypotheses that are prima facie conceivable for us turn out to be impossible, given the right metaphysical analysis of lawhood. A system of laws that mentions the property of being 5 feet away from the Eiffel tower is conceivable, but it might be impossible if the correct account of lawhood implies that laws always relate qualitative properties. Similarly, it may be conceivable that there be laws involving notions with infinite semantic complexity, but this turns out to be impossible if my iterative account is correct. Since the idealized agent that I am envisaging is not afflicted by uncertainty concerning real definitions (including that of lawhood), we can assume that she assigns probability zero to any nomic hypotheses that they rule out.

The nomic hypotheses that are 'live' for this agent impose a privileged partition on the space of possible worlds: two worlds belong to the same cell if and only if they have the same laws. How should our idealized agent distribute prior probabilities over the cells of this nomic partition? Plausibly, in a way that favors simple nomic structures over more complex ones: the probability of each possible nomic hypothesis should reflect its degree of simplicity.[37] Given a traditional (non-iterative) account of laws, the simplicity of some system of laws is just its syntactic simplicity; given an iterative version of the best system account, the simplicity of the lawbook would also depend on its semantic simplicity.

Supposing that there are further epistemic constraints that fix, for every proposition and possible nomic hypothesis, the appropriate prior probability for that proposition conditional on that

[36] Weatherson (2012) and Hawthorne and Dorr (2013) discuss and dismiss a couple of other proposals for connecting naturalness and induction, one that is based on the idea that a property's degree of naturalness correlates with its 'projectability', and one that ties the reasonableness of a prior to its degree of naturalness.

[37] One natural way to implement this idea mathematically is to assign each nomic hypothesis L_i probability $2^{-c}/n_c$, where c is the complexity of the laws according to L_i and n_c is the number of possible nomic hypotheses of complexity c.

nomic hypothesis,[38] we can think of each nomic hypothesis as corresponding to a 'nomic probability distribution'— the distribution that an ideally rational agent would obtain by conditionalizing her prior on the nomic hypothesis. In this setting, the prior of our idealized agent would then be a weighted average of probability distributions associated with possible nomic hypotheses, with the weights determined by the simplicity of those nomic hypotheses. This implements a suggestion by Hawthorne and Dorr (2013): that reasonable priors are weighted averages of all sufficiently 'natural' probability functions, with more natural probability functions weighted more heavily. (But, instead of presupposing a degreed notion of naturalness that covers probability functions, I am suggesting a way to specify the relevant weights in terms of the simplicity of associated nomic hypotheses.)[39]

Let us now see how our idealized agent (a hereafter) would handle Goodman's case. Suppose that $a's$ evidence at t consists entirely of observations of green emeralds at times before the year 3000; the proposition that a needs to conditionalize on would be, roughly, that a has observed n emeralds before the year 3000, and they have all been green. Now consider the two competing hypotheses 'All emeralds are green' (*Green* for short), and 'All emeralds are grue' (*Grue* for short). Their respective credences, after a has updated her rational prior P on their evidence E will be:

$$P_t\ (Green) =\ P\ (Green|E) = \frac{P\,(E|Green)\,P\,(Green)}{P\,(E)}$$

$$P_t\ (Grue) =\ P\ (Grue|E) = \frac{P\,(E|Grue)\,P\,(Grue)}{P\,(E)}$$

Assuming that $P(E|Green) \approx P(E|Grue)$, it follows that the posterior probability ratio of *Green* to *Grue* is roughy the same as the corresponding prior probability ratio. If it is rational for a to be more confident in *Green* than *Grue*, then this must be because it is rational for a to be more confident in *Green* than *Grue* before getting the evidence.

We can see where this asymmetry comes from in a's case, given the simplicity constraint on priors sketched above. Suppose that $L_1, L_2, \ldots L_n$ are all the live complete nomic hypotheses for a. We can express the prior probabilities of *Green* and *Grue* as follows:

$$P\ (Green) = \sum_{i=1}^{n} P\ (Green|L_i)\,P\,(L_i)$$

$$P\ (Grue) = \sum_{i=1}^{n} P\ (Grue|L_i)\,P\,(L_i)$$

[38] Here I have in mind a chance-to-credence norm perhaps together with a moderate indifference principle.

[39] The idealized inductive method considered here also bears some structural similarities to an influential approach due to Ray Solomonoff (1964). Solomonoff proposes an assignment of probabilities to sequences of data encoded in binary strings, where sequences that can be generated by many simple algorithms get higher probability than sequences that relatively fewer and/or more complex algorithms can generate. In my approach, systems of laws are functioning roughly like algorithms function in Solomonoff induction. The ideal agent above will deem a future observation likely when her evidence and that future observation taken together are made probable by many simple systems of laws. Since Solomonoff's method is framed in syntactic terms, it exhibits a problematic kind of language-dependence: if one uses a gruesome encoding of the evidence (or a gruesome Turing machine to run the algorithms), Solomonoff induction will effectively project gruefied properties. My approach avoids this language-dependence by tying nomic hypotheses to the fundamental properties.

Simply put, *a*'s prior credence in each of these two hypotheses is a weighted average of their probability conditional on each nomic hypothesis—the weights being the prior probabilities of each of those nomic hypothesis. So, on this view, *a*'s rational basis for favoring 'All emeralds are green' over 'All emeralds are grue' will have to be that simple nomic hypotheses systematically tend to make *Green* more likely than *Grue*.

While I can't prove that this is true, it does strike me as a plausible conjecture. Note that, among the nomic hypotheses that get high probability by the above simplicity criterion, many will render *Green* much more probable than *Grue*: for example, those that include axioms like 'All emeralds are green' or 'All emeralds are alike in color', and also those that ascribe all emeralds the same (natural) chemical properties, and include familiar generalizations connecting natural chemical properties and natural colors.

Possible nomic hypotheses that would support an assignment of high probability to 'All emeralds are grue' are comparatively harder to come by and tend to be much more complex. Possible nomic hypotheses according to which *Grue* is itself a law-axiom are going to be rare or non-existent; generalizations mentioning grue will be systematically out-competed by systems involving green and other semantically simple properties. Hypotheses on which each emerald has its color fixed randomly are consistent with *Grue*, but don't render it more likely than *Green*. Nomic hypotheses that treat 'All emeralds are either green and observed before the year 3000, or blue and not observed before the year 3000' as a law are highly complex: even granting that 'observed' is natural, such hypotheses will probably lose simplicity points for their disjunctive structure and for mentioning a particular time. Could 'All emeralds are grue' just happen to be very likely by the lights of some physical theory that has a simple formulation in more fundamental terms? This is doubtful given what we know about how emeralds and colors are grounded in more fundamental terms.

The above considerations make it plausible that, for ideal agents like the one described above, there is an asymmetry between *Green* and *Grue* with regards to their prior probabilities. For such an agent, 'All emeralds are green' is inductively learnable on the basis of *E*, whereas 'All emeralds are grue' is not. But how does this bear on what us mortals can learn via rational induction? We almost certainly fall short of the rationality ideal characterized above: our reasoning skills are limited, and we don't have *a* || *priori* access to the full list of possible systems of laws—in part because we are not omniscient about fundamental structure. Given these limitations, it would be miraculous if our responses to evidence always matched those of ideal agents.

Yet, the priors we rely on may be said to be 'highly reasonable' compared to salient alternatives if they underwrite responses to evidence that, as evidence accumulates, come closer and closer to those of an ideal agent—in part by supporting the discovery of new natural properties.[40] How humans are able to do this approximation is a difficult question that I can't hope to address here. But I suspect that we rely on priors that assume that meaningful primitive concepts tend to be natural, and that there are systematic connections between natural properties (of the kind my account predicts).

What about psychologically different agents for whom *grue*, *bleen*, and other such properties seem more natural a priori? Using similar reasoning policies as ourselves, they would arguably end up favoring different hypotheses than us. Here the naturalness fan needs to invoke an exter-

[40] As Titelbaum (2010) shows, there can be no formal updating algorithm that is guaranteed to converge on the list of natural properties, regardless of its starting point. If, as I believe, the ability to do induction well requires figuring out which properties are natural, then this shows that the ability to do induction well presupposes a good enough a priori grip on the world's structure.

nalist maneuver (Sider, 2011; forthcoming): despite their apparent psychological similarity to us, these agents' doxastic states are epistemically inferior if they are based on priors that encode mistaken assumptions about naturalness. Those who insist on an internalist conception of rationality will feel compelled to say that they are as rational as we are. But, even if we grant that such agents are rational or coherent in a minimal sense, there is no pressure to regard them as epistemically on a par with us.

5.4 | Roles for fundamentality

So far I have described three roles that the notion of naturalness has been invoked to play: in the theory of causation/explanation, in the theory of reference, and in the theory of rational induction. I have argued that the nomic account of naturalness provides a promising framework with which to revisit longstanding philosophical questions surrounding each of these notions: (i) How do certain macro-properties get their causal/explanatory powers? (ii) How do naturalness facts get involved in settling the semantic truths? iii) In what sense does our evidence favor simple generalizations involving certain predicates but not others?

The notion of naturalness has been invoked to play many other roles, which I can't cover here. But we should not expect my notion of naturalness to play all of these roles: I think that there are many contexts where the notion of fundamentality is more relevant. Let me mention a couple of examples.

One context where I would invoke fundamentality is in stating general principles about what is possible. Principles connecting naturalness and possibility come in two varieties: expansion principles and contraction principles. Expansion principles tell us that if some things are possible or actual, some other (systematically related) things are also possible: e.g. 'Any way of distributing natural properties across individuals is possible'. Contraction principles constrain the ways in which possible worlds can differ: e.g. 'Every property supervenes on the natural properties' (i.e. no two worlds which are alike with respect to the distribution of natural properties can differ with respect to other qualitative properties).

The notion of naturalness invoked in these modal principles cannot be the nomic notion. With regards to expansion principles, properties that figure in laws at different levels of description are not freely recombinable. There isn't, for instance, a possible world where my intrinsic microphysical structure is the same, but I'm running a high fever. With regards to contraction, the supervenience principle is supposed to concern a minimal set of properties on which everything else is based (like position, spin, charge…), not the set of all properties in a permissive lawbook. This is no problem for my account of naturalness: the above principles can be formulated in terms of the notion of fundamentality.

Similarly, I would have to invoke fundamentality in order to give an account of 'intrinsicality'. While there are many proposals for defining intrinsicality in terms of naturalness, they typically rely on the assumption that all natural properties are intrinsic (e.g., Lewis (1986), Langton Lewis (1998)). This assumption would likely fail for notions of naturalness as permissive as mine (consider properties like being green, being fit, being true…) Thus, the notion of fundamentality is better suited for a theory of intrinsicality.

It is natural to wonder at this point whether the notion of fundamentality that I'm invoking is just Lewis's notion of perfect naturalness. I glossed over this issue earlier in the paper, formulating Lewis's own view in terms of my preferred notion of fundamentality. However, I suspect that the notions are distinct. For Lewis, it is analytic (and necessary) that perfectly natural properties are more natural than all other properties (hence the term 'perfect naturalness'). But I deny the corresponding claim about fundamentality. In the context of a best system account of lawhood, the nomic account of naturalness allows for the possibility of worlds where some fundamental properties are not natural. A fundamental property, like any other, must earn its place in the lawbook by featuring in powerful generalizations. It has an easier time doing so, since appealing to it carries a low simplicity cost. But my view allows for worlds where fundamental properties fail to be natural, because they don't figure in laws.

ACKNOWLEDGEMENTS

Thanks to Laura Callahan, Sam Carter, Chris Frugé, Sam Elgin, Barry Loewer, Ezra Rubenstein, Jonathan Schaffer, Ted Sider, Isaac Wilhelm, Tyler Brooke-Wilson, the participants of the Rutgers Metaphysics Reading Group, the participants of the Oxford Metaphysics Work-in-Progress Group, and two anonymous referees for helpful comments on earlier drafts.

REFERENCES

Armstrong, D. (1978). Universals and Scientific Realism, Vol. 1: Nominalism and Realism, Vol. 2: A Theory of Universals (Cambridge University Press, New York).

Braddon-Mitchell, D. (2001). Lossy laws. *Noûs*, *35*(2), 260–277. https://doi.org/10.1111/0029-4624.00296

Cartwright, N. (1983). *How the laws of physics lie*. Oxford University Press.

Cartwright, N. (1999). *The dappled world: A study of the boundaries of science*. Cambridge University Press.

Cohen, J. (2003). Color: A functionalist proposal. *Philosophical Studies*, *113*(1), 1–42. https://doi.org/10.1023/A:1023074316190

Dasgupta, S. (2018). Realism and the absence of value. *Philosophical Review*, *127*(3), 279–322. https://doi.org/10.1215/00318108-6718771

Davidson, D. (1967). Causal relations. *Journal of Philosophy*, *64*(21), 691–703. https://doi.org/10.2307/2023853

Davidson, D. (1973). Radical interpretation. *Dialectica*, *27*(3-4), 313–328. https://doi.org/10.1111/j.1746-8361.1973.tb00623.x

Demarest, H. (2017). Powerful properties, powerless laws. In J. D. Jacobs (Ed.), *Causal powers* (pp. 38–53). Oxford, United Kingdom: Oxford University Press.

Dorr, C., & Hawthorne, J. (2013). Naturalness. In K. Bennett & D. Zimmerman (Eds.), *Oxford studies in metaphysics: Volume 8* (p. 1). Oxford University Press.

Dretske, F. I. (1981). *Knowledge and the flow of information*. MIT Press.

Fine, K. (2012). Guide to ground. In F. Correia & B. Schnieder (Eds.), *Metaphysical grounding* (pp. 37–80). Cambridge University Press.

Fodor, J. A. (1974). Special sciences (or: The disunity of science as a working hypothesis). *Synthese*, *28*(2), 97–115. https://doi.org/10.1007/BF00485230

Franklin-Hall, L. R. (2016). High-level explanation and the interventionist's 'variables problem'. *British Journal for the Philosophy of Science*, *67*(2), 553–577. https://doi.org/10.1093/bjps/axu040

Gómez Sánchez, V. (2020). Crystallized regularities. *Journal of Philosophy*, *117*(8), 434–466. https://doi.org/10.5840/jphil2020117827

Goodman, N. (1955). *Fact, fiction, and forecast*. Harvard University Press.

Hicks, M. T., & Schaffer, J. (2017). Derivative properties in fundamental laws. *British Journal for the Philosophy of Science*, 68(2). https://doi.org/10.1093/bjps/axv039

Kripke, S. A. (1982). *Wittgenstein on rules and private language*. Harvard University Press.

Langton, R., & Lewis, D. (1998). Defining "intrinsic." *Philosophy and Phenomenological Research*, *58*(2), 333–345.

Lederman, L., & Teresi, D. (1993) *The God Particle: If the Universe is the Answer, What is the Question*. Houghton Mifflin.

Lewis, D. (1983). New work for a theory of universals. *Australasian Journal of Philosophy*, *61*(4), 343–377. https://doi.org/10.1080/00048408312341131

Lewis, D. K. (1973). *Counterfactuals*. Blackwell.

Lewis, D. K. (1974). Radical interpretation. *Synthese*, *27*(July-August), 331–344. https://doi.org/10.1007/BF00484599

Lewis, D. K. (1984). Putnam's paradox. *Australasian Journal of Philosophy*, *62*(3), 221–236. https://doi.org/10.1080/00048408412340013

Lewis, D. K. (1986). *On the plurality of worlds*. Wiley-Blackwell.

Lewis, D. K. (1994). Humean supervenience debugged. *Mind*, *103*(412), 473–490. https://doi.org/10.1093/mind/103.412.473

Loewer, B. (2021). The package deal account of laws and properties (PDA). *Synthese*, *199*, 1065–1089. https://doi.org/10.1007/s11229-020-02765-2

Maudlin, T. (2004). Causation, counterfactuals, and the third factor. In J. Collins, E. J. Hall, & L. A. Paul (Eds.), *Causation and counterfactuals*. MIT Press.

Mill, J. S. (1843). *A system of logic, ratiocinative and inductive: Being a connected view of the principles of evidence, and the methods of scientific investigation*. Longmans, Green, Reader, Dyer.

Mitchell, S. D. (2000). Dimensions of scientific law. *Philosophy of Science*, *67*(2), 242–265. https://doi.org/10.1086/392774

Mitchell, S. D. (2002). Ceteris paribus — an inadequate representation for biological contingency. *Erkenntnis*, *57*(3), 329–350. https://doi.org/10.1023/A:1021530311109

Oppenheim, P., & Putnam, H. (1958). Unity of science as a working hypothesis. *Minnesota Studies in the Philosophy of Science*, *2*, 3–36.

Paul, L. A., & Hall, N. (2013). *Causation: A user?s guide*. Oxford University Press UK.

Putnam, H. (1981). *Reason, truth and history*. Cambridge University Press.

Ramsey, F. (1978). *Foundations: Essays in philosophy, logic, mathematics and economics*, D. H. Mellor *(ed.)*. Humanities Press.

Schaffer, J. (2004). Two conceptions of sparse properties. *Pacific Philosophical Quarterly*, *85*(1), 92–102. https://doi.org/10.1111/j.1468-0114.2004.00189.x

Schaffer, J. (2013). Metaphysical semantics meets multiple realizability. *Analysis*, *73*(4), 736–751. https://doi.org/10.1093/analys/ant069

Schaffer, J. (2014). Writing the book of the world (review). *Philosophical Review*, *123*(1), 125–129. https://doi.org/10.1215/00318108-2366553

Schrenk, M. (2006). A theory for special science laws. In H. Bohse & S. Walter (Eds.), *Selected papers contributed to the sections of gap.6*. Mentis.

Sider, T. (2011). *Writing the book of the world*. Oxford University Press.

Sider, T. (forthcoming). Dasgupta's detonation. Philosophical Perspectives. Forthcoming.

Solomonoff, R. J. (1964). A formal theory of inductive inference. Part i. *Information and Control*, *7*(1), 1–22. https://doi.org/10.1016/S0019-9958(64)90223-2

Strevens, M. (2008a). *Depth: An account of scientific explanation*. Harvard University Press.

Strevens, M. (2008b). Physically contingent laws and counterfactual support. *Philosophers' Imprint*, *8*, 1–20.

Titelbaum, M. (2010). Not enough there there: Evidence, reasons, and language independence. *Philosophical Perspectives*, *24*(1), 477–528. https://doi.org/10.1111/j.1520-8583.2010.00201.x

Weatherson, B. (2013). The role of naturalness in lewis's theory of meaning. *Journal for the History of Analytical Philosophy*, *1*(10). https://doi.org/10.4148/jhap.v1i10.1620

Williams, J. R. G. (2007). Eligibility and inscrutability. *Philosophical Review*, *116*(3), 361–399. https://doi.org/10.1215/00318108-2007-002

How to cite this article: Gómez Sánchez, V. (2022). Naturalness by Law. *Noûs*, 1–28. https://doi.org/10.1111/nous.12433