Special Science Naturalness

Harjit Bhogal

Abstract

As Lewis showed us, there is a compelling theoretical need for a distinction between natural and unnatural properties. Notably, our scientific theorizing should be, as far as possible, in terms of the natural properties. This paper is about how to make sense of naturalness in the special sciences, in a way that is consistent with physicalism. The core idea of my account is that the natural properties of some particular special science form an *explanatory cluster* — most of the facts about these properties are explained well by other facts about those properties. The resulting account, I argue, has a variety of advantages. As well as getting attractive results about which properties count as natural, it helps make sense of the methodological autonomy of the special sciences from lower-level physics, and it helps understand why we should theorize in terms of the natural properties.

As Lewis [1983], among others, showed us, there is a compelling theoretical need for a distinction between *natural* and *unnatural* properties. Natural properties, it is thought, play a variety of important theoretical roles – to do with laws, induction, explanation, and similarity, amongst others. When building our theories we should, as far as possible, do so using the natural properties and not gerrymandered or gruefied properties.

This literature on natural properties has developed in a way that typically closely associates the natural properties with the properties of fundamental physics. Candidate natural properties are things like spin and charge. Further, the natural properties are generally taken to be metaphysically fundamental, unlike the properties of special sciences, which are metaphysically derivative (see, for example, Dorr and Hawthorne [2013]).

However, there is a similarly compelling theoretical need for a notion of naturalness that applies to the properties of the special sciences. When theorizing about macroeconomics, geology, or population ecology, for example, we should also avoid gerrymandered properties and theorize in (relatively) natural terms.

I'm going to give an account of such special science naturalness. It will be a reductive account. I'll start from facts about fundamental-level naturalness, and intertwined facts – like the fundamental physical laws and physical-level causation – and I'll build an account of special science naturalness in those terms.[1] (Some terminology, I'm going to use the term 'F-natural' (for 'fundamental natural') for the traditional, non-special-science natural properties. Where there is no ambiguity I will use 'natural' to mean 'special science natural' but never to mean F-natural.)

[1] The account is, in the terminology of Gómez Sánchez [2020], a *foundationalist* project.

The account is based on the idea that natural properties in the special sciences form *explanatory clusters* – most of the facts about such properties can be explained well by other facts about such properties. I argue that this account gets attractive results for what the natural properties are; it makes sense of why we should theorize in terms of the natural properties; and it captures important features of our special scientific practice, in particular, the methodological autonomy of the special sciences from lower-level science.

Difficulties, however, do arise with respect to the possibility of permuting these explanatory clusters to generate intuitively unnatural properties. Towards the end of the paper I point to some strategies for dealing with this problem.

1 Alternative Accounts

But we should start by considering a few other possible approaches to special science naturalness. I won't say enough about these approaches to rule them out, just enough to locate and motivate my preferred account.

1.1 Primitivism

As I said, my account will be reductive, it will not take special science naturalness to be primitive. But primitivism is, of course, an option. Fodor [1974] is plausibly interpreted as a primitivist. For example:

> I am suggesting, roughly, that there are special sciences not because of the nature of our epistemic relation to the world, but because of the way the world is put together: not all natural kinds…are, or correspond to, physical natural kinds. (p. 113)

It's easy to hear Fodor as saying that special science natural properties are just part of 'the way the world is put together' – they are basic parts of the world. Or perhaps Fodor's view is not quite primitivism about special science natural properties, but is primitivism about a very closely related entity – special science laws. The natural properties can then be understood as those that are involved in the laws.[2] Either way, there are reasons to be dissatisfied with this approach. Firstly, accepting that there are irreducible special science entities is clearly a denial of physicalism, even if there are some physicalist theses – in particular, token physicalism, not type physicalism – that Fodor's view satisfies (see Loewer, 2009, p. 229). That's enough motivation for me to try a different approach.

Secondly, Loewer [2009, pp. 229–232] and Gómez Sánchez [2023, pp. 102–3] argue that primitivism implies an unpleasant redundancy in explanations of similarity, causal relations, counterfactuals, and so on. Such facts would be explained by facts about fundamental level naturalness and laws and, separately, by special science naturalness and laws. Further, Gómez Sánchez [2023, pp. 101–2] argues that primitivism would involve unattractive violations of *Purity* – the principle that no fundamental truth involves a non-fundamental property [Sider, 2011, section 7.1]. But more detail here is unnecessary – we have more than enough motivation to try a reductive account.

[2] See Gómez Sánchez [2023, section 3.3] for discussion.

1.2 Length of Definition

Perhaps the most commonly discussed reductive approach starts by defining a graded notion of F-naturalness and claiming that the special science natural properties are highly (or, at least, highly enough) F-natural. The standard account of graded F-naturalness is that a property is more F-natural if it has a shorter definition in terms of a *fundamental language* – one where all the predicates stand for perfectly F-natural properties and relations (Lewis [1986b, p. 61]; Sider [2011, section 7.3]).

There are a few reasons to be doubtful of this kind of approach. Sider [1995, pp. 363–4] and Hawthorne [2007, pp. 433–4] both argue that many special science properties, both intuitively natural and intuitively unnatural ones, may have infinitely long definitions in terms of the natural properties, and so would all count as equally natural to each other.

Further, even if special science properties have finite definitions, it is not at all clear that the intuitively natural properties have shorter definitions than the intuitively unnatural ones. Consider, for example, the economic property of inflation. The definition of this property would be incredibly long and we have very little idea about how it would go. The thought that our actual concept of inflation will turn out to have a relatively simple definition compared to intuitively unnatural economic properties, and similarly for other intuitively natural special science properties, just seems hopelessly optimistic. Gómez Sánchez [2023, section 2.2] goes further – she argues that we can take the simplest definition of some high-level property, like *pain* or *inflation*, and we can just delete some part of that definition. (If the definition involves disjunctions of lower-level realizers we can just delete a disjunct. If it's defined in terms of chains of functional definitions we can just delete some part of the characterization of the relevant function.) But doing so does not yield a property that is intuitively more natural.

What's more, it is highly implausible that when we judge certain special science properties to be natural and others to be unnatural what we are responsive to is the length of the definition of the property in terms of the properties of fundamental physics. Again, we have almost no idea how such a definition would go. And it's hard to see why an economist, for example, would care about length of definition, even if they did know. The length of definition strategy doesn't integrate well with our scientific practice and our judgements of naturalness.

1.3 Gómez Sánchez's Iterated BSA

The most well-developed and plausible account of special science naturalness is from Gómez Sánchez [2023]. In one way her account is very simple – the natural properties are the ones that are involved in scientific laws. But in another way the account is very intricate. For her account to yield special science natural properties she needs an account of special science laws – one which does not already build in a distinction between the natural and unnatural properties of the special sciences.

The account she gives is an ingenious variation of the classic Best System account of laws (BSA) [e.g. Lewis, 1983, p. 42–3]. On the classic view, we start with the Humean mosaic, the distribution of occurrent, non-modal, events that occur throughout spacetime.[3] We then systematize all of the facts about the mosaic in a way

[3]More precisely, we can take the Humean Mosaic to be constituted by the intrinsic physical state of each spacetime point (or each pointlike object) and the spatio-temporal relations between those points.

that best balances simplicity and informativeness. The core idea of the BSA is that the laws are the axioms of the system that best balances simplicity and informativeness. A huge amount of detail is suppressed here – filling this idea out is rather complicated in ways that we shouldn't get into now.[4]

Importantly, the classic BSA restricts the language that systems can be formulated in. The motivation for this is to avoid the *predicate F problem*. Consider a predicate, F, that is instantiated by all and only the things that exist in the actual world. A system with only one axiom, '$\forall x F(x)$', would be extremely simple and (in the sense of informativeness that Lewis endorsed), maximally informative because the truth of $\forall x F(x)$ rules out all non-actual worlds. But, clearly this should not count as a law of nature. Lewis's solution is to say that predicates like F are disallowed – systems must be formulated in a language where the basic predicates of the language express perfectly F-natural properties and relations.

Gómez Sánchez needs to relax this restriction without falling prey to the predicate F problem, and without building in anything that would make the account of special science naturalness circular. The key innovation is to *iterate* the systemization procedure. The first stage of this iterative process is very similar to the classic BSA, with one notable change. Systems are not restricted to being formulated in terms of a language where the basic predicates of the language express perfectly F-natural properties and relations (a language she calls *FL*). Rather, an additional dimension of simplicity – she calls it *semantic complexity* – is added on which systems score badly if they use predicates that have a large *FL-complexity*, where the FL-complexity of a predicate is given by that predicate's length of definition in terms of FL. So, the predicate F would score very badly, because it's extremely hard to define in FL, thus avoiding $\forall x F(x)$ counting as a law. The output of this first stage is a set of axioms that count as laws, and a set of properties involved in those laws that thereby count as natural.

We then run this procedure again, but with the adaption that what matters for simplicity is not the length of the definition in language FL, but the language RL_1, where this language has basic predicates for all of the perfectly F-natural properties *and for all of the properties that count as natural by virtue of being part of axioms of the best system at the first stage.* The output, again, is a set of axioms that count as laws, and a set of properties involved in the laws that count as natural.

Then we run the same procedure again, but the relevant language is RL_2, where this language has basic predicates for all of the perfectly F-natural properties and for all of the properties that count as natural by virtue of being part of axioms of the best system at the *first or second stage*. Repeat this infinitely. An axiom that is part of the best system at any level is a law, and a property that is involved in any of those axioms is natural.

The thought is that this iterative procedure can make room for the distinctive properties of the special sciences, properties that are massively multiply realizable, since the properties of some higher-level science, macroeconomics or oceanography, for example, could be functionally defined in terms of the properties of some lower-level science, which could be functionally defined in terms of an even lower-level science, and so on down to the fundamental. This chaining of functional definitions can be captured by Gómez Sánchez's iterative procedure where systems build upon each other. That's the hope at least.

[4] See, for example, Braddon-Mitchell [2001], Schrenk [2006], Cohen and Callender [2009], Loewer [2012], Hall [2015], Loewer [2024], Bhogal and Perry [2017], Hicks [2018], Dorst [2018], Jaag and Loew [2020] among many others.

The procedure may also be able to explain the way in which *green* is more natural than *grue*. Since, plausibly, the functional definition of green in terms of the winners of lower-level competitions is simpler than the functional definition of grue – "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" [Gómez Sánchez, 2023, p. 116].

There is a huge amount to say about the details of this account. There are are so many choice points and variations of the classic BSA – and the adaptions Gómez Sánchez makes multiply the complexity. What's more, she describes her account merely as an 'existence proof' that there could be an account of lawhood that meets relevant desiderata for extracting an account of special science naturalness. An iterative BSA procedure, she suggests, can be developed in very different ways.[5] So, I'm not going to nitpick about the details of her BSA procedure.

Instead I'm going to focus on a couple of big picture issues, in order to motivate an alternative view. Firstly, on Gómez Sánchez's account the higher-level natural properties have a certain lack of *autonomy* from the lower-level physics. This is because the account has a *bottom-up* structure. The natural properties of economics, for example, are constructed, brick by brick, from the fundamental physical properties. My concern is that there may be too much dependence of the special science natural properties on the details of the fundamental physics. If the details of the fundamental physics were different, then that could radically change what higher-level properties count as natural.

Of course, sometimes this is the right result. If the laws of physics were such that there were no regularities connecting demand for a good and the price of that good then plausibly the natural properties of economics would be different, or perhaps would not exist at all. But if we restrict our attention to situations where the social world looks roughly as it does in the actual world, then it doesn't seem like facts about what is going on at much lower scientific levels should matter for economic theorizing.

For example, if there happens to be an extremely simple, elegant theory of everything given in fundamental terms, then such a theory may well win the best system competition at every stage (this is especially possible if semantic complexity weighs highly against a system). But even if this were so there would, it seems, still be value in economic theorizing and there would still be natural economic properties.[6] More generally, it's rather unclear why, for various ways that the fundamental physics might be, we should be confident that laws about mid-level dry goods would win the best system competition at any level. For example, if the fundamental physical laws depicted the world as extremely holistic, with states of the world understood as extremely non-local, then it's hard to be confident that running the BSA again can take us to the kind of local, non-holistic laws that we get in, for example, various parts of chemistry and biology.

[5] For example, she suggests, following other work of hers [Gómez Sánchez, 2020], that we might think of laws as summarizing sets of physically nearby worlds, and not the mosaic of events.

[6] What would the value of economic theorizing be in a world where there is a simple theory of everything? Partly, it would be pragmatic – even if the laws of the theory of everything were simple applying them to extremely complicated economic systems which are constituted by ridiculously large numbers of fundamental entities would likely be beyond our capabilities. But, what's more, it's common to accept that there are non-pragmatic advantages to giving explanations in higher-level terms, rather than lower-level ones, though there is controversy about precisely what this advantage consists in. Perhaps the higher-level explanation properly identifies the *difference-makers* [e.g. Strevens, 2008, Woodward, 2010], perhaps it is *proportional* to the explanandum [e.g. Yablo, 1992, Woodward, 2018], perhaps it is more *modally informative* [Jackson and Pettit, 1992], or perhaps it is more *robust* [Wilson, 1994, Weslake, 2010].

The general thought here is that the practice of higher-level sciences is largely autonomous of the lower scientific levels. As long as the social world works roughly as it does then the practice of economics would, and should, remain the same. But this threatens tension with Gómez Sánchez's bottom-up approach. In fact, there is reason to think that high-level sciences, like economics, are largely autonomous of even other higher-level sciences. If the economic property of a *preference* turned out to be hard to define in lower-level scientific terms, like those of psychology or neuroscience, then that would likely not substantially affect the practice of economics. In fact, this arguably is the actual situation. One way of understanding (part of) the *neuroeconomic* critique of traditional economics is that it's hard to make sense of the concepts that economists use to analyze, say, human decision-making, in lower-level neuroscientific terms (see, for example, Camerer et al. [2004, pp. 572-3], Glimcher [2011, chapter 1]), but economists have typically asserted the autonomy of their metholodology (most influentially, Gul and Pesendorfer [2008]).

There is a lot of structure to Gómez Sánchez's account that she might appeal to or vary, in a bid to capture more of this autonomy of the higher-level sciences. However, there is a sense in which this autonomy problem is only a dramatization of the second problem that I want to discuss.

Gómez Sánchez's account faces a kind of *missing value problem*. Why does naturalness have the normative power that it does? As we noted at the start of the paper, we should theorize in terms of the natural properties. But why? Dasgupta [2018] raises this as an objection to primitive accounts of F-naturalness. Primitivists, he argues, need to accept a kind of irreducible normativity. Dasgupta claims that this is very unattractive. I think reasonable people can differ on this point – arguably the primitivist can happily accept that it's a basic epistemic principle that there is value to theorizing in terms of the natural properties (see Sider [2011, especially sections 1.1 and 4.5]). But this is not a plausible option for the reductivist, like Gómez Sánchez. There isn't a primitive normative principle that we should theorize in terms of the properties outputted by the iterative BSA procedure (Gómez Sánchez accepts this kind of point in footnote 4.). Rather, an explanation must be given for why we should theorize in terms of those properties.

But, it's hard for me to see what this explanation could be. What epistemic or other advantage comes from the economist theorizing in terms of the properties outputted by the iterative BSA procedure? Remember, we start with a system that best balances simplicity, informativeness about the whole mosaic and accuracy, where simplicity includes semantic simplicity – how the properties of the system are defined in terms of a language where the basic predicates refer to perfectly F-natural properties – then we iterate this process, in the way we described above, adapting what makes for semantic simplicity by adding the properties already judged to be natural to the relevant language. It's rather hard for me to see why economist's theorizing should be sensitive to what properties arise at, say, the 14th level of this process. Why, for example, are the explanations that an economist gives worse if they appeal to properties that don't show up in this hierarchy? Even if the account can be tuned to get intuitive results about the extension of special science naturalness it not clear how it can be cleanly integrated with our special science practice.

This is only a very superficial discussion of Gómez Sánchez's account. But still, I take the doubts about the bottom-up nature of the account to motivate looking at an alternative approach.

2 Motivating the View

In this section, I'll motivate my approach before stating the view a bit more precisely in the next section.

Consider the relation between naturalness and *explanation*. Special science naturalness is needed to rule out explanations that seem inappropriately gerrymandered. (See, for example, Kitcher [1981, section 8], Strevens [2008, section 3.6.1], Franklin-Hall [2016] and Weslake [2015].)

Take an economic fact e about the price of a particular good rising. Assuming that the fundamental laws are deterministic then there are certain physical states of the world $b_1, \ldots b_n$ such that e holds if and only if one of those states held at time t. And so consider a property P that the world possesses if and only if it is in one of the states $b_1, \ldots b_n$. Now consider the putative explanation of e from the fact that the world has property P at time t. This is not a good explanation of e. In fact, it doesn't really look like an explanation at all.

Clearly, P shouldn't be used in explanations – it's an unnatural property. If we can find what is wrong with explanations involving P, independent of any thoughts about naturalness, that should illuminate the unnaturalness of P and other properties.

So, what is wrong with the explanation of e from the fact that the world has property P at t? Here is an intuitive thought: the explanation is bad because P is gerrymandered. Talk of 'gerrymandered properties' is common in discussions of naturalness, but it is not always clear what it means. When I say P is gerrymandered, I'm not saying that P is overly complicated – it's not clear in what sense P is more complicated than any other high-level property. And I'm not using 'gerrymandered' as a synonym for 'unnatural' – clearly doing that would not advance the discussion.

Rather, to speak loosely for a moment, what makes P gerrymandered is that it is 'designed' solely in order to get one, very specific, result – to be part of an explanation of the fact e – and further it seems to only have this one use. Subsequently, the result that P is designed to achieve seems, in some sense, cheap.

Consider the electoral district rezoning that spawned the word 'gerrymandering'. In that case, the borders of electoral districts are drawn in order to achieve very specific electoral ends. For example, a party might pack most of the supporters of their opposition into a few districts so they lose those districts by very large amounts but win the majority of districts by much smaller amounts. In paradigm cases the only reason anyone would draw the boundaries in that way is to achieve this very specific electoral goal.

P, I suggest, is gerrymandered in a similar way. It's 'designed' solely to play a role in the putative explanation of e from the fact that the world is P at time t in a way that seems inappropriate. This intuition is forceful, but talk of properties being 'designed' for certain ends is rather imprecise. Don't we always design our explanations to fit the phenomenon under investigation? And in what sense do we design properties? So we need to be a bit more careful here. I'm going to discuss two ways in which P seems to be gerrymandered where typical special science properties are not.

2.1 Portability

P is not portable – it is not useful in explaining a wide range of explananda. If we are interested in explaining *e*, or some very closely related facts, then *P* might have a role to play. But for explaining other economic facts – for example, a fact about the marginal cost of producing good *G* – appealing to *P* is a bad strategy. And we wouldn't appeal to *P* to explain biological, chemical, or sociological facts either. In the vast majority of cases *P* is not useful for explaining what we want to explain.

Genuine scientific practice typically involves explaining a wide range of facts by appealing to the same set of properties. For example, it was considered a deep benefit of Newtonian mechanics that it could explain such a wide range of phenomena, both celestial and terrestrial, starting from a sparse base of properties – *force, mass, velocity*, and so on [Kitcher, 1981, section 3]. Similarly, it's a good feature of standard microeconomics that it can explain so much in terms of a few basic properties – properties like *agent, goods, preference, demand, supply, factor of production, choice, price*. We do not find non-portable properties like *P* being used in genuine explanatory practice.

And there is a good reason for this, because appealing to non-portable properties to explain leave us lacking a type of understanding. Consider an extreme case, where we explain every different fact by appealing to a different non-portable property. Fact *e* is explained by the world having *P* at *t*; fact *f* by the world having *Q* at t_1; fact *g* by the world having *R* at t_2; and so on. Such explanations seem unenlightening.

I don't have a full account of the nature of scientific understanding to appeal to (though I'll say a tiny bit more in section 2.3) but I suspect that the reason they seem unenlightening is because the collection of such explanations portrays the world as unstructured. The above explanations don't reveal the connections between *e*, *f* and *g*; we don't understand how these explananda and the corresponding explanations relate to each other. Regardless of this hypothesis, though, the key thought is just that there is something suboptimal about explanations involving non-portable properties.

2.2 Connections to other Properties

This suggests a second way in which *P* seems gerrymandered – it lacks substantial connections to other properties. *P* is not interestingly connected to the properties involved in *e*, the fact about good *G* rising in price. The property *P* is not generally related to price rises, for example. Although *P* is closely connected to one particular instance of prices rising – the world being *P* at *t* picks out all and only the worlds where *e* holds – it is not so connected to other instances.

More generally, some sets of properties have rich and substantial explanatory connections to each other – the Newtonian and microeconomic cases illustrate this. Force, mass and velocity are deeply explanatorily connected to each other – facts about velocity are very often explained by facts about force and mass, for example. Similarly, the microeconomic properties are very closely explanatorily connected. Facts about demand for goods are explained by agents' preferences; facts about certain preferences are explained by other preferences; facts about certain choices are explained by preferences and prices; facts about prices are explained by facts about demand

and supply; facts about the existence of certain goods are explained by the demand for other goods; facts about the supply of goods are explained by the demand for certain factors of production; facts about the demand for factors of production are explained by the price of the goods that they are used in producing; and so on.

The property P does not seem to have such rich and substantial connections to any such set of properties. P seems somewhat *isolated* from other properties.[7]

2.3 Explanatory Networks

As we have seen, when we explain a fact we shouldn't use a property, like P, that seems to be cooked up specifically and only for that explanatory purpose. Rather, our practice favors properties that have a much richer set of explanatory connections – properties that we can use to explain a variety of facts, and which have close explanatory connections to other properties. Our practice, that is, favors properties that are embedded in *rich explanatory networks*.

As we noted earlier, there is a thought about understanding here: we seem to derive a certain understanding from theorizing in terms of properties that are deeply connected to each other and are part of a rich explanatory network.

Consider, for example, price changes in various goods. One good rose in price because of increased demand. Another increased in price because of decreased supply. A third good's price remained stable because while demand increased, supply also increased in a compensating way. A fourth good's price fell because although demand increased, supply also increased by a very large amount.

Seeing the connections between these facts, and how they are explained in related ways, greatly helps our understanding of those facts and of price changes more generally. In fact, it is precisely by illustrating the relations between such cases that introductory economics textbooks elicit understanding. We can see the connections between these facts only because we, in explaining them, appealed to properties like demand, supply and price that are portable to various explanatory contexts and are closely explanatorily related to each other – that is, by appealing to properties that are part of a rich explanatory network.

Again, I don't have an account of understanding to give here, but previous discussions of understanding do fit with these thoughts. For example, Grimm [2012, p. 103], in his survey of the literature, characterizes the notion of understanding discussed by epistemologists as 'roughly, the good of being able to "grasp" or "see" how the various parts of the world were systematically related'. The explanations of the price rises in terms of demand and supply help give us this type of understanding, while explaining using non-portable properties does not allow us to see these systematic connections.

Similarly, Elgin [2007, p. 35] claims that 'To understand the Athenian victory involves more than knowing the various truths that belong to a suitably tethered comprehensive, coherent account of the matter. The understander must also grasp how the various truths relate to each other and to other elements of the account.' Isolated explanations, like the explanation of e from P, don't yield understanding, rather we need to be able to

[7] You might worry whether P really is isolated in this way, maybe there is some set of, perhaps unnatural, properties that P has substantial explanatory connections to. This point will be discussed in section XX.

see the connections between the relevant properties. Similar thoughts are common in the literature on scientific understanding.

It's highly intuitive, I think, that elucidating these systematic connections between properties gives us an explanatory benefit. Isolated explanations give less understanding than a unified body of explanations. The moral of this discussion is that the properties we should use in theorizing – the natural properties – are part of rich explanatory networks. In the next section I'll give an account of naturalness based on this idea.

3 The View

The guiding idea, from the last section, is that perfectly special science natural properties are part of rich explanatory networks. In this section I will develop this idea – giving an account of what makes for rich explanatory networks and how they connect to naturalness.

But just before that we should deal with an objection to the approach.

3.1 A Circularity Concern

The view I'm developing accounts for naturalness in terms of explanation – natural properties are part of rich explanatory networks. But there is a naturalness constraint on explanations – we can't use extremely unnatural properties, like the property *P*, in explanations. But if naturalness is a prerequisite for explanation then giving a reductive account of naturalness in terms of explanation seems to be off the table.

The solution is to ignore, so to speak, the naturalness constraint on explanation when giving an account of naturalness. More specifically, consider *explanation**. Explanation* is just the relation of explanation but with the constraint that only natural properties can be used to explain removed. An explanation* would also be an explanation if the properties used to do the explaining* were natural. We should give the account of naturalness in terms of explanation* not explanation. So, natural properties are members of rich explanatory* networks.

Of course, this idea of 'subtracting' the naturalness constraint is easier to implement for some accounts of explanation than others. In this paper, I'm going to be working with an account of explanatory correctness inspired by Lewis [1986a]. The account is a very minimal one, it just says that *A* explains *B* when *A* gives information about the nexus of causal relations that led up to *B* and where *A* doesn't contain very unnatural properties and is not very disjunctive. It is trivial to generate explanation* from this – just delete the clause about naturalness.

As I'll discuss more in section 3.2.1, although I'm working with this simple account of explanatory correctness my account isn't tied to it. I could appeal to other accounts of explanation too, as long as I remove the naturalness constraint. Some accounts require a little more work to factor out the component that appeals to special science naturalness. For example, causal accounts of explanation may build in special science naturalness by having a naturalness constraint on causation. In this case we would need to extract the naturalness constraint at the point of the account of causation.

Perhaps there are accounts of explanation where special science naturalness is so deeply enmeshed that there is no way to extract it and be left with anything coherent. I would not be able to use those accounts to give a reductive account of special science naturalness. This, then, is a commitment of my approach. For the rest of the paper, I'm going to work with the minimal causal account of explanation and the corresponding account of explanation*.

3.2 Explanatory Clusters

We are now in a position to give the account of naturalness.

Consider the concept of explanatory* clustering. Explanatory* clustering is a feature of a set of properties. It is a graded notion – the explanatory* clustering of a set of properties Γ increases with increases in:

1. The proportion of the facts about properties in Γ that are explained* by other facts about properties in Γ
2. The goodness of these explanations*
3. The number of facts about properties in Γ

So, a set of properties scores very highly on explanatory* clustering if *most* of the facts about properties in Γ are explained* *well* by other facts about Γ and there are a *large number* of facts about Γ.

The basic idea is that a property is perfectly natural if and only if it's part of a sufficiently explanatorily* clustered set of properties. It turns out we need a couple of tweaks on this basic idea, so the account of natural properties is as follows:

A property is perfectly natural if and only if (i) it's part of a sufficiently explanatorily* clustered set of properties, (ii) none of the properties in the set are definable in terms of the other properties in the set and (iii) that set is a local maximum of clustering – that is, we can't increase the explanatory* clustering by making small changes to the properties in the set.

The second condition is required because otherwise my account would imply that if *A* and *B* are perfectly natural then *A*&*B* is perfectly natural. The third condition is required otherwise we could take a set of properties that is well over the required threshold for explanatory* clustering and then add an arbitrary property and the resulting set would still likely be over the threshold, making that arbitrary property count as natural.

To simplify the terminology let's call a set of properties that meets conditions (i)-(iii) an *explanatory* cluster*. So, a property is perfectly natural if and only if it is part of an explanatory* cluster.

Importantly, that this implies that a property is perfectly natural if and only if it is part of an explanatory cluster (note the lack of an asterisk on 'explanatory'). This is because if an explanatory* cluster is made up of natural properties, then it is an explanatory cluster and if something is an explanatory cluster then it is also an explanatory* cluster because the conditions on explanatory* clustering are strictly weaker than those on explanatory clustering.

If a set of properties is an explanatory cluster then the members will have the characteristics of being portable and having substantial connections to other properties. Imagine a set of properties $\{A, B, C, D\}$ that form an explanatory cluster. Then there are a large number of facts about those properties and most of them can be explained well in terms of facts about *A*, *B*, *C* or *D*. This means that *A*, *B*, *C* and *D* are portable – they are useful in explaining a wide range of facts – facts about *A*, *B*, *C* or *D*.

Similarly, the properties *A*, *B*, *C* and *D* are tightly explanatorily connected to each other. There are many explanatory connections between the facts about these properties. In fact, although this is not guaranteed by anything said so far, if large numbers of facts about *A*, *B*, *C* or *D* are explained by other facts about those properties, then it is likely that there are some systematic explanatory connections between these properties. That is, it is likely that there are fairly uniform, perhaps lawlike, ways in which certain combinations of instantiations of the properties *A*, *B*, *C* and *D*, explain other instantiations of the properties. If this were not the case then it would be extremely coincidental that most of the facts about those properties were explained by other facts about those properties.

This is primarily an account of perfectly special science natural properties. But we can extend this by defining a graded notion of special science naturalness – using the length of definition of a property in terms of the perfectly special science natural properties.

3.2.1 Explanatory Goodness

The account of special science naturalness appeals to explanatory goodness. (Strictly it appeals to explanatory* goodness, but I'll come back to this soon.) An account of explanatory goodness tells us when explanations are better or worse. What's to come will not require us to fix on one particular account of explanatory goodness. Rather, a loose intuitive grip on explanations being better or worse is all that we need.

But for definiteness, and to help us get an intuitive grip on explanatory goodness, I will briefly outline an account of explanatory goodness that we might use. My preferred account is based on the common idea that the explanans should be *proportional* to the explanandum. The explanans shouldn't be far more specific than the explanandum – it shouldn't be the case for example, that the explanans contains a huge amount of physical detail but the explanandum is a very abstract fact. But neither should it be the case that the explanans is very abstract while the explanandum is very specific. A variety of authors (e.g. Yablo [1992], Woodward [2018], Weslake [2010], among others) have been driven by similar thoughts.

The account I will appeal to, that of Bhogal [2021], develops this idea in a rather simple way. It starts by assuming some account of explanation and then builds explanatory goodness out of this. For simplicity, let's continue working with the minimal causal account of explanation. This says that *A* explains *B* when *A* gives information about the nexus of causal relations that led up to *B*. But, as we have discussed, there has to be a naturalness constraint too – the explanans cannot contain very unnatural properties or be very disjunctive.

Given this, we can outline two dimensions of explanatory goodness. The first dimension of explanatory goodness, precision, says that explanations are better if in more of the physically possible worlds where the explanans is true, the explanans explains the explanandum. For example, the explanation of an ice cube melting from the fact that it was dropped in warm water scores very highly, but not maximally, on precision because nearly all the microstates that could realize the system are such that dropping the ice cube in warm water does lead to the melting. The explanation of the glass of water spilling because I knocked the table on which it was sitting with my knee scores worse on precision, because there are many physically possible worlds where I knock the table with my knee but that does not lead to any spilling (for example, those where I don't knock the table hard enough).

The second dimension, ROBUSTNESS, says that explanations are better if in more of the physically possible worlds where the explanandum is true, the explanans explains the explanandum. An explanation of the extinction of a species which cites the introduction of a predator will score higher on ROBUSTNESS than the explanation of the same fact which cites the fundamental laws and the fact f which describes the total state of the world at a particular time t. That is because more of the physically possible worlds in which the species goes extinct are such that the extinction is explained by the predator's introduction than explained by f. f cannot explain the extinction in many worlds because it holds in only very few physically possible worlds.

ROBUSTNESS tells us that the explanans shouldn't hold in few possible worlds when the explanandum holds in many. PRECISION tells us that the explanans shouldn't hold in many worlds when the explanandum holds in few. Together they tell us that there should be a proportionality between explanans and explanandum.

Of course, there is a lot more to say about this approach to explanatory goodness. But this is just an example of one type of approach to explanatory goodness that could fit into the account of naturalness I'm developing here. However, strictly speaking my account of naturalness appeals to explanatory* goodness – a sense of goodness that applies to explanations* not explanations. But it is easy to adapt the account of explanatory goodness we just outlined into one of explanatory* goodness. Just remove the naturalness constraint from explanation, thus yielding an account of explanation*, and the same definitions of PRECISION and ROBUSTNESS will give us an account of explanatory* goodness.

3.3 Autonomy and Missing Value

The account of naturalness I'm developing here has, I suggest, the promise of avoiding the issues raised for Gómez Sánchez's account.

Firstly, the account is built to make sense of the autonomy of the special sciences. If a set of properties Γ is an explanatory cluster then most of the facts about Γ are well explained by other facts about Γ. We do not need to look outside Γ, so to speak, in order to give good explanations. Γ, therefore, defines a subject matter that is particularly amenable to autonomous investigation.

What's more, the account is 'top-down' in a way that contrasts with Gómez Sánchez's account – what matters for naturalness is the connection between higher-level properties, not how those properties are built from the lower-level properties. As long as there are relevantly robust connections between the higher-level properties that will make those properties natural, regardless of what is going on at, for example, a microphysical level. So, if the social world looks roughly as it does then the properties of microeconomics will be natural, even if the microphysics is very different.

Further, as I've noted, the account gives us resources for explaining the value of theorizing in terms of the natural properties. I've suggested that theorizing in terms of explanatory clusters generates more understanding than giving isolated explanations. And, what's more, there is an obvious practical advantage for intellectually limited inquirers like ourselves to theorize in terms of a small set of properties, rather than appealing to completely different properties every time we are looking to explain a different fact.

What we need to do now is see whether this account gets plausible results for what the natural properties are.

3.4 Letting in Natural Properties

Does the account line up with our judgments about naturalness in the practice of the special sciences? Answering this question fully is a huge project, it involves detailed study of the causal and explanatory relations of various special sciences. Right now, I'm just going to outline some general considerations which should lead to optimism that my account gets attractive results in the majority of cases.

Let's take this in two steps. In this subsection, consider whether the account allows in the properties we want; that is, whether it classifies the properties used in paradigm special science explanations as natural. In the next subsection, consider whether the account rules out unnatural properties – like the property *P*.

Consider microeconomics and a set of properties that are fairly distinctive to that field, for example, *agent, goods, preference, demand, supply, factor of production, choice, price*. This set will count as highly explanatorily clustered because these properties are explanatorily interconnected in a large range of ways. Facts about demand for goods are explained by agents' preferences; facts about certain preferences are explained by other preferences; facts about prices are explained by facts about demand and supply and so on. In fact, standard microeconomic theorizing tells us about these connections in very systematic ways; there is a large body of theory regarding exactly how a few facts about these properties explain the others. It looks like, then, these microeconomic properties are part of an explanatory cluster and so count as natural.

Or consider classical genetics. *Gene, allele, trait, dominance and inheritance* are all closely connected by good explanations and will form a cluster. Further, consider population ecology and properties like *population, generation, predator, prey, carrying capacity*; and thermodynamics and properties like *temperature, pressure, entropy*.

But we can do more than just survey examples. It's plausible that this account of naturalness will allow in paradigm special science properties – the ones that are part of good explanatory theories. A theory of a particular special science will be framed in terms of a set of properties Γ with axioms about those properties and rules for inferring other propositions about those properties. If the theory is successful it will be able to explain well many facts about Γ. In this case, the properties of Γ would form an explanatory cluster because some facts about Γ would explain well a large number of other facts about Γ. So, when we have special sciences, like microeconomics and classical genetics, for which we have good explanatory theories then the basic properties of those theories will likely form an explanatory cluster.

But there are special sciences for which we do not possess such a successful explanatory theory. In such cases we may have good reason to believe that a theory could be developed in the future using roughly the properties that the special science used now. This would give us reason to think that that those properties are natural. For special sciences where we have little reason to think that such an explanatory theory can be developed, we have little reason to think that the basic properties of those special sciences form explanatory clusters, and so little reason to think that they are natural. This is an acceptable result. There's no reason to think that it's easy for us to find out what these natural properties are, and no reason to think that current scientific practice always maps on to the natural properties in the world.

3.5 Ruling out Unnatural Properties

So the account gets reasonable results in classifying properties actually at work in the special sciences as natural. Does the account rule out the right properties? Or can we have explanatory* clusters that contain intuitively unnatural properties? (Call these 'unnatural clusters' for short.) I'll discuss some reasons for optimism before considering a tricky issue about permuting properties.

Let's start by seeing why it is not easy to construct unnatural clusters. Explanatory* clustering involves a kind of 'closure' that makes it hard to add unnatural properties to a cluster. Imagine that we add an unnatural property, like P, to a set of natural looking properties, say the standard set of microeconomic properties, call this Δ.

Facts about P explain* well one particular fact about Δ – the fact *e*. But this is not enough for $\{\Delta \cup P\}$ to be more clustered than Δ because facts about Δ and P together do not explain* well any of the other facts about P. For example, the fact that the world does not possess property P at time t_4 is not well explained* by the properties in $\{\Delta \cup P\}$. So $\{\Delta \cup P\}$ scores worse than Δ on clustering and so cannot be an explanatory* cluster because it is not a local maximum of clustering.

Adding gerrymandered properties like P to traditional special science properties will not generally increase clustering because the special science properties won't explain* well most of the facts about P. For a property to count as natural it really needs to be deeply integrated with a cluster of other properties – it's not enough to just have the relatively minimal connections that P does to the microeconomic properties.

If we do manage to cook up a property Q that fits closely with the traditional economic properties – one that explains* well lots of the facts about the traditional economic properties and where many of the traditional economic properties explain* well facts about Q – then we have just discovered a new natural economic property. If there really are these rich and deep connections between Q and the other economic properties then Q should be integrated into our economic theorizing in the way that other natural economic properties are.

So adding a property to some traditional special science properties won't cause a problem for my account. Could we instead build an explanatory* cluster where all of the properties are intuitively unnatural? The first thing to note is that there is no reason to think that we have discovered all the special sciences. If we find a novel set of properties that form a rich explanatory* network, we should be open to accepting that those properties are natural and that we have discovered a new special science, even if those properties might initially strike us as unnatural.

3.5.1 Permuting Clusters

But there are strategies for developing novel explanatory clusters where we are disinclined to think of the resulting properties as natural. In particular, if we can take an explanatory* cluster and transform each property in it such that the facts about those transformed properties have the same causal relations as the original properties, and these causal relations hold in the same way across physically possible worlds, then this transformed set will also be an explanatory* cluster. I'll end by discussing this problem and a couple of potential responses – though what I say will be too brief to be fully satisfying.

The most obvious transformation involves 'property switching' – of the kind exhibited by the classic definition of grue. Let grue be green before the year 2000 and blue after. And let bleen be blue before 2000 and green after. We can think of grue and bleen as 'switching roles' in the year 2000, previously grue played the green-role, after it plays the blue-role, and similarly with bleen. Of course, this idea can be extended to more than just two properties. So if we have an explanatory* cluster of properties, *P*, *Q*, *R*, *S* and *T*, we can form a transformed set of properties, *A*, *B*, *C*, *D* and *E* where each of the transformed set 'switches roles' at a specific time. So, for example, *A* plays the *P*-role up to 2000 and the *Q*-role after, and so on for the other transformed properties. If done correctly, it looks like the transformed set of properties may also be an explanatory* cluster. This is because the causal relations between facts about the transformed set will mirror the causal relations between facts about the original set of properties, in both this and other physically possible worlds. If this is right, then we will get a large equivalence class of clusters that are transformations of each other in such ways.

There are a couple of approaches we might take to this problem. I'll tentatively mention a more ambitious approach before pointing to a less ambitious one. I'm undecided which is the right approach to take.

Notice that the original cluster and the transformed cluster have different characteristics that we might appeal to in an effort to favor one over the other. For example, certain regularities will be more simply expressible in terms of the original cluster of intuitively natural properties than in terms of the gruefied, transformed properties. For example, there are, in fact, regularities about demand and price, and such regularities hold before and after the year 2000. It will be more complicated to express such regularities in terms of gruefied properties that switch roles in the year 2000 than in terms of the intuitively natural properties.

We might develop this thought into a way of distinguishing between clusters of properties that are transformations of each other. Consider the various clusters and see which set of properties are best for formulating simple, informative, regularities about the world. The suspicion is that the transformed properties will do a less good job at this than the intuitively natural properties. The economic patterns do not, in fact, change in the year 2000, so expressing them using properties that switch roles then will be needlessly complicated.

The strategy here is very much in the spirit of Loewer's [1996, 2024] package deal account of natural properties. The core idea is to run the BSA account of laws, minus the restriction that systems have to be formulated using only predicates that refer to natural properties. Then we take the laws that win the competition – the axioms that are simple and informative – and say that the properties involved in those laws count as natural. The major challenge that this approach has to deal with is, as mentioned in section 1.3, the predicate F problem – what's to stop a law like $\forall x F(x)$, winning the competition? Loewer, of course, develops strategies for dealing with this issue. And we can understand Gómez Sánchez's account as an adaption of Loewer's that aims to avoid this problem.

The strategy I'm suggesting is very similar – we consider which cluster of properties systematizes better. But the predicate F problem is much less pressing, since I'm only using the mechanics of systemization to choose between explanatory* clusters that are transformations of each other. It's much harder to see, then, how a predicate F-style strategy could trivialize the account since we don't get free rein in devising problematic properties.

This, then, is one strategy for dealing with the problem of transformed clusters. I'm somewhat undecided about it though – I'm unsure whether this strategy will work for every way we might transform explanatory* clusters. So, I'll end by pointing to a less ambitious approach we might take too, though this approach is consistent with the more ambitious one we have been describing.

Notice that there may well be other ways we can identify valuable features of scientific theorizing that would motivate ruling out certain transformations of explanatory* clusters. For example, while I've been stressing that special sciences are, in practice, largely autonomous of each other, they are not completely autonomous. There is interdisciplinary work, and it is valuable. So, one thing we might want is for the various explanatory clusters of different sciences to cohere with each other, at least to some degree. So, we might implement a kind of 'metaclustering' condition, which would give us a principled reason to rule out certain transformed clusters.

But once we have done as much as we can in identifying the valuable features of scientific theorizing and noting how they restrict the range of properties that we should use in such theorizing, we might still not have ruled out all of the intuitively unnatural properties. At this point, the question arises, are people who theorize in terms of those intuitively unnatural properties making any sort of mistake?

As we noted in section 1.3, if we are primitivists about naturalness we might plausibly think the answer is yes – there is a way that the world is put together and communities that theorize in terms of other properties are in violation of a basic epistemic principle.

But, this is much less plausible if we are not primitivists. In this case, if the identifiable values of special science theorizing are exhausted without picking out a unique set of properties that we should theorize in terms of then we should accept that special sciences naturalness isn't as restrictive as we might have thought. In this context, I'm inclined to think that there might be a range of properties, some of which seem fairly strange and gruefied to us, such that a community that theorizes in terms of those properties isn't doing anything wrong. And so, we might come to terms with an account of special sciences naturalness that doesn't rule out all of the properties that seem strange to us.

References

Harjit Bhogal. Induction and the glue of the world. *Australasian journal of philosophy*, 99(2):319–333, 2021.

Harjit Bhogal and Zee R Perry. What the humean should say about entanglement. *Noûs*, 51(1):74–94, January 2017.

David Braddon-Mitchell. Lossy laws. *Nous*, 35(2):260–277, June 2001.

Colin F Camerer, George Loewenstein, and Drazen Prelec. Neuroeconomics: Why economics needs brains. *The Scandinavian journal of economics*, 106(3):555–579, October 2004.

Jonathan Cohen and Craig Callender. A better best system account of lawhood. *Philosophical Studies*, 145(1): 1–34, 2009.

Shamik Dasgupta. Realism and the absence of value. *The Philosophical review*, 127(3):279–322, January 2018.

C Dorr and J Hawthorne. Naturalness. *Oxford studies in metaphysics*, 8, January 2013.

Chris Dorst. Towards a best predictive system account of laws of nature. *The British journal for the philosophy of science*, 70(3):877–900, January 2018.

Catherine Elgin. Understanding and the facts. *Philosophical studies*, 132(1):33–42, January 2007.

Jerry Fodor. Special sciences (or: The disunity of science as a working hypothesis). *Synthese*, 28(2):97–115, October 1974.

Laura Franklin-Hall. High-level explanation and the interventionist's 'variables problem'. *The British Journal for the Philosophy of Science*, 67(2):553–577, June 2016.

Paul W Glimcher. *Foundations of neuroeconomic analysis*. OUP USA, January 2011.

Stephen R Grimm. The value of understanding. *Philosophy Compass*, 7(2):103–117, January 2012.

Faruk Gul and Wolfgang Pesendorfer. The case for mindless economics. *The foundations of Positive and normative Economics: A handbook*, 1, January 2008.

Verónica Gómez Sánchez. Crystallized regularities. *The Journal of Philosophy*, 117(8):434–466, 2020.

Verónica Gómez Sánchez. Naturalness by law. *Nous*, 2023.

Ned Hall. Humean reductionism about laws of nature. In Barry Loewer and Jonathan Schaffer, editors, *A Companion to David Lewis*, volume 61, pages 262–277. John Wiley & Sons, Ltd, Oxford, UK, April 2015.

John Hawthorne. Craziness and metasemantics. *The Philosophical review*, 116:427–440, January 2007.

Michael Townsen Hicks. Dynamic humeanism. *The British journal for the philosophy of science*, 69(4):983–1007, December 2018.

Siegfried Jaag and Christian Loew. Making best systems best for us. *Synthese*, 197(6):2525–2550, June 2020.

Frank Jackson and Philip Pettit. In defense of explanatory ecumenicalism. *Economics and philosophy*, 8(1):1–21, January 1992.

Philip Kitcher. Explanatory unification. *Philosophy of science*, 48(4):507–531, 1981.

David Lewis. New work for a theory of universals. *Australasian Journal of Philosophy*, 61(4):343–377, December 1983.

David Lewis. Causal explanation. In David Lewis, editor, *Philosophical Papers Vol. Ii*, pages 214–240. Oxford University Press, January 1986a.

David Lewis. *On the Plurality of Worlds*. Basil Blackwell, London, January 1986b.

Barry Loewer. Humean supervenience. *Philosophical Topics*, 24(1):101–127, January 1996.

Barry Loewer. Why is there anything except physics? *Synthese*, 170(2):217–233, September 2009.

Barry Loewer. Two accounts of laws and time. *Philosophical studies*, 160(1):115–137, January 2012.

Barry Loewer. *Laws of Nature and Chances: What Breathes Fire into the Equations*. Oxford University Press, 2024.

Markus Schrenk. A theory for special science laws. In H Bohse and S Walter, editors, *Selected Papers Contributed to the Sections of GAP.6*. mentis, January 2006.

Theodore Sider. Sparseness, immanence, and naturalness. *Nous*, 29(3):360–377, 1995.

Theodore Sider. *Writing the Book of the World*. Oxford University Press, January 2011.

Michael Strevens. *Depth: An Account of Scientific Explanation*. An Account of Scientific Explanation. Harvard University Press, 2008.

Brad Weslake. Explanatory depth*. *Philosophy of science*, 77(2):273–294, 2010.

Brad Weslake. The problem of disjunctive explanations. January 2015.

Robert A Wilson. Causal depth, theoretical appropriateness, and individualism in psychology. *Philosophy of science*, 61(1):55–75, 1994.

James Woodward. Causation in biology: Stability, specificity, and the choice of levels of explanation. *Biology and Philosophy*, 25(3):287–318, January 2010.

James Woodward. Explanatory autonomy: the role of proportionality, stability, and conditional irrelevance. *Synthese*, pages 1–29, October 2018.

Stephen Yablo. Mental causation. *The Philosophical review*, 101(2):245–280, 1992.