Crash Course on Naturalness*

Theodore Sider February 17, 2024

Contents

1. The distinction

Consider:

*Further reading: Dorr (2019); Dorr and Hawthorne (2013); Sider (1993, 2011).

Being a deer	being either a deer or a chair or a coin in my pocket
Having 5kg mass	being identical to a_1 or to a_2 or to … or to a_n
being green	being grue
being five feet from	causing or being a prime number of feet from

(An object is "grue" if and only if it is either green and first observed before 3000AD, or blue and not first observed before 3000AD (Goodman, 1955*b*).)

There is clearly some sort of difference between the properties on the left and the properties on the right. The ones on the right seem arbitrary, or cooked-up. Their members don't seem to share anything in common. They don't seem to be of interest in science.

The properties and relations on the left are sometimes said to "carve nature at its joints", or to be "natural kinds" (although 'kind' doesn't seem like the right word when relations are concerned—note the last item on the left). David Lewis, who was an important defender of the importance of the distinction (1983*b*; 1984; 1986*a*, pp. 59–69), used the term 'natural properties and relations'.

Let's look at some questions about naturalness.[1]

2. Its significance

> "For as I bear [the notion of naturalness] in mind considering various topics in philosophy, I notice time and again that it offers solutions to my problems." Lewis (1983*b*)

In the early 1980s, Lewis became convinced by his friend David Armstrong that the distinction between natural and non-natural properties was a legitimate one to appeal to in philosophical theorizing. From that point on, he repeatedly relied on it. The quotation describes a feeling that will be familiar to fans of naturalness. If you bear naturalness in mind, you too will notice how it always solves your problems. Naturalness really does seem indispensable in metaphysics. To my mind, it is this, more than "intuitions" about examples, that is the most compelling argument for accepting and relying on the distinction.

[1]Another important issue, which I won't discuss, is whether there is just *one* important notion of naturalness or whether there are many. See Dorr and Hawthorne (2013).

Lewis (1983*b*) talks about some of these "solutions to his problems"; see also Lewis (1986*a*, pp. 59–69).[2] It's a good idea to look at a few of them quickly, since they give the feel for what naturalness is supposed to be.

2.1 Fundamentality

Lewis doesn't explicitly cite this as one of the "problems" that naturalness solves, but to my mind it's right at the front of the list. Consider this quotation:

> Physics has its short list of 'fundamental physical properties': the charges and masses of particles, also their so-called 'spins' and 'colours' and 'flavours', and maybe a few more that have yet to be discovered. In other worlds where physics is different, there will be instances of different fundamental physical properties, alien to this world... And in unphysicalistic worlds, the distribution of fundamental physical properties won't give a complete qualitative characterisation of things, because some of the 'fundamental' properties of things will not be in any sense physical. What physics has undertaken, whether or not ours is a world where the undertaking will succeed, is an inventory of the sparse properties of this-worldly things. Else the project makes no sense. It would be quixotic to take inventory of the abundant properties—the list would not be short, nor would we discover it by experimental and theoretical investigation. (Lewis, 1986*a*, p. 60)

It's natural to think of physics as being, at least in part, an inquiry into what "the most basic bits of the world" are. But simply listing the *things* that are, in fact, the "most basic bits"—all the subatomic particles and points of spacetime, perhaps, wouldn't suffice. Physicists are also trying to discover their most basic *features*: charge, mass, etc. (And also, the laws governing those features.) But what are the "most basic", or "fundamental" properties and relations? Answer: the natural properties and relations

(Or maybe physics isn't interested in *all* of the natural properties and relations, just the ones that help explain matter in motion.)

I hate to use the term 'analytic', but if you're looking for anything that is analytically built into the notion of naturalness, this is it: the perfectly natural properties are the most fundamental properties.

It might not seem very contentful, though, to say that natural properties are

[2] And also Sider (2011).

"fundamental". Is there any upshot of this idea that doesn't contain 'fundamental' or a synonym? Here are three possibilities:

Aim of physics It's (part of) the aim of physics to discover the natural properties and relations

Completeness Everything [depends on/reduces to/supervenes on] the distribution of natural properties and relations

Minimality There is no proper subset of the natural properties and relations on which everything [depends on/…]

(I myself think of minimality as an epistemic principle. Thus I would replace minimality with some claim that, other things being equal, theories are more choiceworthy when they posit fewer natural properties and relations (Sider, 2011, p. 219).)

2.2 Duplication and intrinsicality

Consider these two closely related concepts: duplication and intrinsicality. Duplicates are things that are exactly alike. Intrinsic properties are properties whose instantiation by an object is purely a matter of what that object is like, taken in and of itself—i.e., without regard to what is going on outside that object.

Is there any way to define these notions?

How about "objects are duplicates iff they have the same properties"? No good; even duplicates don't share *all* of their properties in common. Two duplicate electrons might differ over the property *being within five feet of some proton*, or the property *being identical to e* (where '*e*' is the name of some particular electron).

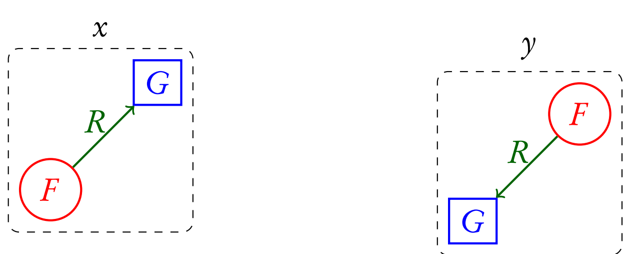
Let's try defining 'intrinsic' instead. How about: "a property is intrinsic iff it could be had by an object that is alone in the universe"? No good: as Lewis (1983*a*) pointed out, the property *being alone in the universe* isn't intrinsic, but it satisfies the definition.

Lewis pointed out that we could define 'duplicate' in terms of 'intrinsic': duplicates are objects with exactly the same intrinsic properties. Or we could define 'intrinsic' in terms of duplicates: intrinsic properties are those that can never differ between any pair of duplicates (whether from the same or different possible worlds). But we're going in a circle.[3]

If your attempts to define something keep meeting counterexamples, this can be a sign that you need bigger guns—that you need to use more powerful notions in the definiens.[4] That's what Lewis did; his bigger gun was 'natural'. Using 'natural' he was able to define 'duplicate'; and then he gave the aforementioned definition of 'intrinsic' in terms of 'duplicate'. Here was his definition of 'duplicate':

> … two things are duplicates iff (1) they have exactly the same perfectly natural properties, and (2) their parts can be put into correspondence in such a way that corresponding parts have exactly the same perfectly natural properties, and stand in the same perfectly natural relations. (Lewis, 1986*a*, p. 61)

Intuitive idea: duplicates are things composed of exactly ultimate parts, put together in the same way:



x and y are duplicates

The remainder of this section is a digression!

In addition to using the notion of naturalness, Lewis's definition of 'duplicate' also uses 'part', which Lewis doesn't define in other terms. Why was he ok with

[3] Also, the latter definition is contentious because it treats necessarily coextensive properties as having the same intrinsicality status, and also fails to count identity properties as being intrinsic. See Marshall (2023, section 2).

[4] 'Natural' wouldn't be "more powerful" if it could be defined in terms of 'intrinsic' or 'duplicate'; then it would just be another member of the circle. But it doesn't seem possible to define 'natural' in terms of those notions.

that?

It's hard to say exactly why, though it's clear that he did think of mereology—the theory of parts and wholes—as being privileged in various ways.[5] He thought that mereology—both its notions and its assumptions—was a legitimate "starting point" for philosophical theorizing: it is completely clear, nonvague, beyond reproach in any way, etc.

But what if you weren't willing to take parthood for granted? I doubt that you could then break into the "intrinsicality-duplication" circle from the outside.[6] But here's a question: could you begin with intrinsicality/duplication and define parthood? In particular, could you define the parts of x as those objects that affect x's intrinsic properties?

The best attempt at turning this idea into a definition, I think, is this:[7]

> The whole-part relation is the most inclusive two-place relation R with this feature: if a property, P, is intrinsic, then so is the property *being R-related to something that has P*

But notice that this an "implicit" definition. It provides a way of uniquely picking out the parthood relation out of a domain of relations that "already" contains the very relation we are "defining". Explicit definitions, on the other hand, provide us with a replacement expression for the one being defined, which can be seen to be in good standing without making assumptions about what properties and relations there are. Thus whereas explicit definitions provide a

[5] It's hard to find an explicit statement of his view here, though see Lewis (1991, chapter 3).

[6] You *can* break into a related circle, consisting of notions that are restricted to pluralities of simples. You could say that simpleses xx and yy are "duplicateses" iff they can be put into one-to-one correspondence such that corresponding objects have the same perfectly natural properties and relations, and say that a plural property of simpleses is intrinsic iff it never differs between duplicateses. What can't be broken into, I'm saying, is the circle of notions of intrinsicality and duplication as applied to individual entities which may be composite.

[7] The idea is to turn "inheritance principles" of intrinsicality (Sider, 2007) into a definition. Actually, the definition should be a little more complicated, using the idea of intrinsic *relations* (Lewis, 1986*a*, p. 62): the whole-part relation is the most inclusive relation R such that for any $m+n$-place intrinsic relation R', the following relation is also intrinsic:

$$\lambda y, x_1, \ldots, x_n . \exists y_1, \ldots, y_m \big(R(y, y_1) \wedge \cdots \wedge R(y, y_m) \wedge R'(y_1, \ldots, y_m, x_1, \ldots, x_n)\big)$$

If we think of R as whole-part, the displayed relation can be glossed thus: "has parts that bear R' to". (The definition in the text is the special case where $n = 0$ and $m = 1$.)

way of avoiding recognizing the property or relation being defined as perfectly natural, implicit definitions don't. And indeed, although I won't go into this in detail, I think that even given the definition (and even if duplication was a natural relation), I think there would remain pressure to regard parthood as a natural relation.[8]

2.3 Laws of nature

A law of nature, to put it intuitively but vaguely (and somewhat contentiously), is a general statement describing how objects behave, which is necessarily true in some sense of 'necessity' that is supplied by "nature" (as opposed to whatever grounds "metaphysical" necessity). Newton, for example, thought that it was a law of nature that "An object at rest remains at rest, or if in motion, remains in motion at a constant velocity unless acted on by a net external force."

What is it to be a law of nature? As with any metaphysical question, "primitivism" is always an available answer: some statements are laws of nature, others aren't, and that's all that can be said. But Lewis was in general opposed to primitivism about all kinds of necessity; thus he was in the market for a reduction of lawhood.

One traditional reductive account of laws of nature is this:

Regularity theory A law is nothing more than a universally true regularity—a sentence of the form "All Fs are Gs" that is true at all times and places, where F and G are "suitable" predicates.

But it soon became clear that this just wouldn't work.[9] "Every solid lump of gold is less than 1 million pounds" isn't a law (even if it is true); "Every solid lump of Uranium 235 is less than 1 million pounds" *is* a law (let's suppose); but it's hard to think of a definition of 'suitable' that would exclude one but not the other.

Lewis defended instead a very influential "best-system theory" of lawhood:[10]

> I adopt as a working hypothesis a theory of lawhood held by F. P. Ramsey in 1928: that laws are "consequences of those propositions which we

[8]Assuming that we believe in composite objects at all; see Sider (2013) for an argument against composite objects based on similar considerations.

[9]Armstrong (1983, Part 1) is a good survey.

[10]See Lewis (1973, pp. 73–4; 1983*b*, pp. 366–8; 1986*b*, pp. 121–4; 1994). See also Beebee (2000); Hall (2015); Loewer (1996).

> should take as axioms if we knew everything and organized it as simply as possible in a deductive system". We need not state Ramsey's theory as a counterfactual about omniscience. Whatever we may or may not ever come to know, there exist (as abstract objects) innumerable true deductive systems: deductively closed, axiomatizable sets of true sentences. Of these true deductive systems, some can be axiomatized more *simply* than others. Also, some of them have more *strength*, or *information content*, than others. The virtues of simplicity and strength tend to conflict. Simplicity without strength can be had from pure logic, strength without simplicity from (the deductive closure of) an almanac… a contingent generalization is a *law of nature* if and only if it appears as a theorem (or axiom) in each of the true deductive systems that achieves a best combination of simplicity and strength. (Lewis, *Counterfactuals*, p. 73.)

Or anyway, that is how Lewis originally stated the view. But later, he pointed out a problem:

> Different ways to express the same content, using different vocabulary, will differ in simplicity… In fact, the content of any system whatever may be formulated very simply indeed. Given system S, let F be a predicate that applies to all and only things at worlds where S holds. Take F as primitive, and axiomatise S (or an equivalent thereof) by the single axiom $\forall x F x$. If utter simplicity is so easily attained, the ideal theory may as well be as strong as possible. Simplicity and strength needn't be traded off. Then the ideal theory will include (its simple axiom will strictly imply) all truths, and *a fortiori* all regularities. Then, after all, every regularity will be a law. That must be wrong. (Lewis, 1983*b*, p. 367)

Lewis's solution: require that the language in which the theory is stated must contain predicates only for perfectly natural properties and relations.

2.4 Reference magnetism

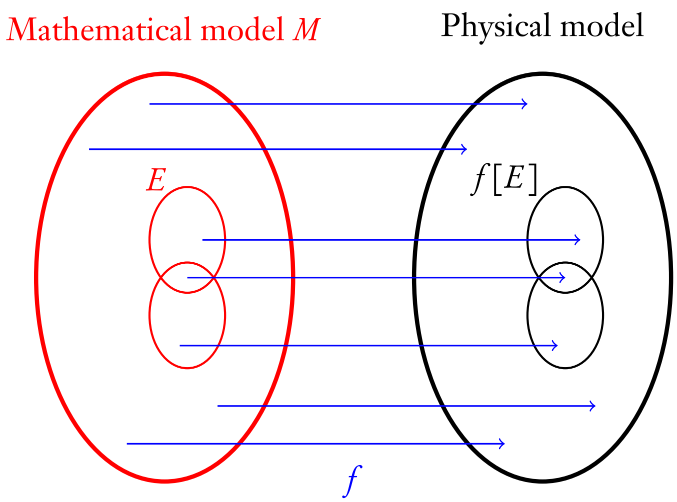
How do our words acquire their meaning? What glues a word to its referent? This question is sometimes called the question of "metasemantics" (although that term is sometimes used in other ways). It's the question of the metaphysics of reference.

When we introduce a new predicate, we often give its meaning in terms of other words we already understand. We might, for instance, define 'vixen' as meaning 'female fox', if we already understood 'female' and 'fox'. Even if we don't give necessary and sufficient conditions, we might use words we already understand to *constrain* the meaning of the new word—to give a necessary condition for a property to be meant by it. If we already understand 'electron', we might say, for example, that the word 'charge' must be so understood that the sentence 'every electron is charged' comes out true.

But this cannot be a *general* metaphysical account of meaning, since it leaves unanswered the question of how the words we "already understand" get their meanings in the first place.

Here is one—very simplistic, and very incomplete—idea of how words in general get their meanings. For each predicate, there is a meaning-giving theory for that predicate: a bunch of sentences used to constrain its meaning. (These meaning-giving theories can overlap; we aren't assuming any sort of "foundationalism" here.) Put all the meaning-giving theories together into one big theory, T. All the predicates have the meanings that they need to have, in order for T to be true. Better: any interpretation (i.e., assignment of meanings to predicates) in which every member of T is true is a *correct* interpretation (i.e., a correct description of what we really mean).

Hilary Putnam (1978, part IV; 1980; 1981, chapter 2) pointed out that given this sort of metasemantics, it follows that it's basically inevitable that everything we believe is in fact true. (Instead of taking that as a reductio of the metasemantics, he spun it as an objection to realism. See Lewis (1984) for criticism.) Let P be the set of sentences we believe about physics, say. Suppose that P is consistent with T, in the sense that there exists some mathematical model (domain together with assignment of predicate extensions) M in which all the sentences in P and all the sentences in T are true. And suppose further that there exists some such M whose domain is the same size as the set of physical objects—that is, that there is some one-to-one function f from M's domain onto the set of physical objects. Then if we construct the "images" of the predicate extensions in M through the function f, we get a model whose domain is the set of physical objects, in which exactly the same sentences are true as in M:



The diagram shows how extensions get sent to their image extensions; e.g., E to $f[E]$. Relations like inclusion and overlapping are preserved by this process, as the diagram depicts. This is why if you begin with a model and image the extensions through some one-to-one function, exactly the same sentences are true in the resulting model.

Since the same sentences are true in the physical model as in M, every sentence in both T and P is true in the physical model. If the metasemantics is correct, the physical model is a correct one. Thus (to put it roughly), the mere *consistency* of P with the meaning-giving theory implies that P is true given some correct interpretation.[11]

Lewis said: since this consequence is absurd, something else beyond making-the-meaning-giving-theory-true must be required of correct interpretations. What might that be? Lewis's answer was naturalness. The correct assignment of meanings must, other things being equal, assign reasonably natural properties as predicate-meanings.[12]

[11] Why did we need the physical model; why not just directly cite the mathematical model as a correct interpretation in which P is true? Because one might refrain, in some way, from taking talk of the abstract model at face value, while taking the physical model at face value.

[12] I've described how Lewis incorporated naturalness into a metasemantics based on the idea that "the meaning-giving theory determines reference". Lewis didn't in fact accept that sort of metasemantic account; he just used it as a simple illustration of how naturalness is needed to supplement metasemantics.

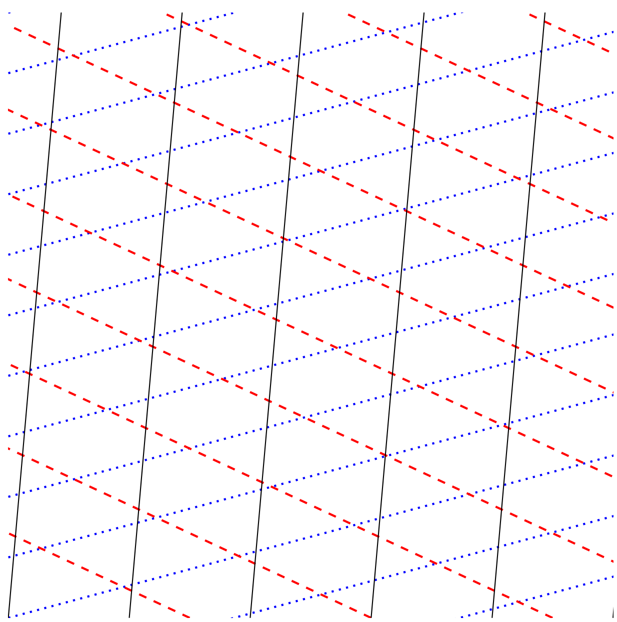
(There's a lot more to say here…)

2.5 Structure of space and time

In physics, it's common to speak of what the "structure" of a space is. For example, we ordinarily assume that space is "isotropic": there are no "distinguished" directions. For instance, there is no notion of "up" that is "intrinsic to space", or "built into space", or "part of space's structure". 'Up' just means: away from the center of mass of the nearest really massive thing (and thus refers to different directions depending on where the utterer is located). Directions, that is, are not part of space's distinguished structure.

What *is* intrinsic to space? *Distance*, according to many people (but not everyone). Unlike directions, there are facts about distance (such as that a certain pair of points are exactly as far from each other as a certain other pair of points) that are "built into space itself".

But what does all this mean? A direction, let's say, is just a maximal set of nonintersecting lines. There are all sorts of directions—infinitely many in fact. This diagram depicts three:



What does it mean to say that none of these direction is "intrinsic to" or "built into" space, or that none of them is "distinguished"?

And there are many functions from points to real numbers with the right formal features to count as "distance functions"—infinitely many of them, in fact. Many of these, intuitively, do *not* correspond to the "intrinsic" distance facts. For instance, suppose d is the function that assigns the "real" or "correct" or "built-in" distances between any pair of points. Where f is any one-to-one function from the set of points onto itself, we can define another function d' from d as follows:

$$d'(x,y) = d\big(f(x), f(y)\big)$$

This function will have the same formal features as d does, so that it counts as having the "right formal features" if d does, but it may assign very different values. If d counts points x and y as being exactly as far from each other as points z and w are (that is, if $d(x,y) = d(z,w)$), d' need not agree (maybe $d'(x,y) \neq d'(z,w)$).

So: what does it mean to say that distances are "built into the structure of space", but that directions are not? An obvious answer is that no particular direction is definable from the perfectly natural properties and relations of space—there are no perfectly natural relations such as "x is north of y"—whereas a particular distance function *is* definable from the perfectly natural properties and relations over points (or, perhaps, a set of distance functions is definable, whose members differ only over the unit of measure). Perhaps, for example, the relation *x is exactly as far from y as z is from w* is perfectly natural.[13]

3. Regimentation

What kinds of things, exactly, are, or are not, natural? Relatedly, how do we regiment talk of naturalness?

According to Lewis, "abundant" properties and relations include both natural and nonnatural properties and relations. For him, to any set of possible individuals whatsoever, there exists an abundant property that is had by all and only those individuals. (Indeed, Lewis *defined* abundant properties as sets of possible individuals.) Thus the abundant properties and relations are closed under logical operations, such as conjunction, disjunction, and (in a sense) negation.

[13] See Wallace (2019) for a very different way of thinking about the structure of space; see Sider (2020*b*, chapter 5) for a discussion of some related issues.

Ignoring issues having to do with semantic and set-theoretic paradoxes, any meaningful predicate picks out an abundant property or relation. Given this setup, Lewis then said that naturalness is a distinction amongst the abundant properties and relations. Some of them are natural, and some are not. Thus 'natural' is a predicate of abundant properties and relations.

What are the alternatives?

One alternative comes from David Armstrong (1978). For him, the only properties and relations that exist are natural properties and relations. Let's reserve his word 'universals' for properties thus understood. Thus there is no distinction between natural and unnatural universals.

Suppose you don't believe in properties and relations at all—you're a nominalist. You might then speak of *predicates* (which are linguistic entities) as being natural. The predicate 'is a deer' is a *natural predicate*; the predicate 'being either a deer or a chair or a coin in my pocket' isn't. A disadvantage is that we would be limited to evaluating for naturalness those predicates we have in our language.

Yet another alternative is to speak of naturalness in a higher-order language.[14] In a language allowing quantification into predicate position, and allowing second-order predicates, one could formulate the assertion that "there exists at least one natural property" as follows: $\exists F \mathscr{N}(F)$, where $\mathscr{N}$—a second-order predicate—symbolizes "is natural".[15]

A related question to that of regimentation is the scope of naturalness. Does it make sense to ask, of an individual thing, whether it is a natural entity? And does it make sense to ask whether logical operations, such as disjunction, quantification; or borderline logical operations such as necessity and possibility, are natural?

The question of scope is related to that of regimentation because if we *can* speak

[14] See Sider (2020*a*) for an introduction to higher-order languages.

[15] Second-order predicates are different from predicates *of* predicates, which is what the previous paragraph discussed. A predicate of predicates is a predicate that attaches to a name of a predicate to form a sentence, as in 'The predicate 'being a deer' is natural'. A second-order predicate is an expression that attaches to a (first-order) predicate—not a name of that predicate, but rather, the predicate itself—to form a sentence. Sentences formed using predicates of predicates are *about words*; sentences formed using second-order predicates are (in general) not.

Sider (2011, section 6.3) uneasily defends a somewhat related view, but which is not coupled with an acceptance of higher-order quantification, and which is not formalized.

of naturalness in these cases, then we'll need a regimentation under which this makes sense. For example, Lewis regiments 'natural' as a first-order predicate; but such predicates must attach to singular terms, which refer to entities. So disjunction and existential quantification would need to be treated as entities.[16] Shifting to an appropriate higher-order language, however, would avoid this problem (Dorr, 2019, section 3.4).

Another related question is whether the canonical locution for talking about naturalness is comparative. In Lewis's regimentation, for example, do we say 'abundant property p is more natural than abundant property q'? Or simply: 'abundant property p is natural simpliciter'?[17] Lewis himself spoke of naturalness as coming in degrees, but treated perfect naturalness as primary since he defined degrees of naturalness in terms of it: a property or relation's degree of naturalness is a function of how simply and shortly it can be defined in terms of the perfectly natural properties and relations.[18] Notice that some vagueness/lack of objectivity can creep into the definition of relative naturalness ("simply and shortly"), but that there is no corresponding vagueness or lack of objectivity in perfect naturalness.

An alternative to Lewis would be to take 'as-or-more natural than' as undefined, and define a perfectly natural property and relation as one that is as-or-more natural than any other property or relation. That strikes me as a step in the wrong direction. Consider the property of *being money* and the property of *being a rock*. Which is more natural? Or are the two equally natural? It's hard to believe that there are always objective (or nonvague, etc.) answers to such questions. Whereas it's comparatively easy (I think) to believe that there are always objective, nonvague answers to the question of whether a given property is *perfectly* natural. So I think it's best to separate perfect naturalness from relative naturalness.

4. Definition

How might we define (reduce, analyze) 'natural'?

[16] See Sider (2011, chapter 6) on this issue.

[17] See Lewis (1986*a*, p. 61); Schaffer (2004); Sider (2011, section 7.11).

[18] Verónica Gómez Sánchez (2021), like Lewis, begins with perfect naturalness and then defines degrees of naturalness; but she has a quite different, and better, I think, definition.

4.1 Primitivism

Some people *don't* reduce it. They simply use 'natural' without defining it—"take it as a primitive notion".

As Lewis pointed out (Beebee and Fisher, 2020, letter 352), this doesn't mean that you think the notion *can't* be defined. You might simply refrain from defining it, for various reasons—e.g., you don't know of a good definition, you want to remain neutral on the definition, etc.

A stronger view of this sort would be that naturalness itself is "primitive" in some sense—that it has some metaphysical status makes it impossible or inappropriate to define (in some sense of 'define') the term 'natural'. Here is a statement of a view of this sort:

> Perfect naturalness is perfectly natural. Therefore there is no "metaphysical definition" of perfect naturalness, since in such a definition, the definiendum is always less natural than the definiens.

However, this raises a number of difficult issues:

1. 'Perfect naturalness is perfectly natural' needs to make sense. One's regimentation of naturalness would need to allow this.
2. It would need to be true. This would rule out a very appealing "physicalist" picture: in creating the world, God only needed to specify the basic physical stuff, and not also which properties and relations are natural.
3. This might not pass muster at USC.

Regarding 3: it's natural to think that in any true definition, the definiendum is *identical* to the definiens, understood in a higher-order sense. Still, one might say, the definiendum must be less natural than each *component* of the definiens. What this might mean exactly would take some time to chase down, but whatever it means, it's likely to collide with certain views about grain. Given some views about grain, for instance, the notion of a component of a property isn't in good standing. And if I metaphysically define a property p as the conjunction of some property q with itself, the principle seems to imply that p is less natural than q, whereas given coarse grain, p will be identical to q.

4.2 Defining naturalness in terms of laws

If one wants to define 'natural', what are the possibilities?

We've already mentioned Armstrong on universals. If you believe in abundant properties and relations in addition to universals, you can define a natural property or relation as one that corresponds to some universal.

Here's another important possibility: define a natural property or relation as one that figures in some *law of nature*. Let's think about that for a bit.

Lots of people who think about these topics are playing a certain game, which we might call the "game of directed definition". The object of the game is to define as many philosophically important notions as you can, from an appropriate starting point of undefined notions. What count as appropriate undefined notions is disputed, but participants in the game generally agree that the fewer undefined notions, the better. But what isn't disputed—what is built into the rules of the game—is that definitions can't go in circles.

As we saw, Lewis himself defined 'law' in terms of 'natural'. And his philosophical conduct makes it very clear that he was playing the game of directed definition. So he could not accept a definition of 'natural' in terms of 'law of nature'. Indeed, 'natural' was at the very bottom of his edifice of definitions—it occurred in practically all of his philosophical definitions (starting in the early '80s, when he converted to team naturalness). The fact that 'natural' is capable of playing this role is a big part of what attracted Lewis to it.

Lewis's rejection of the definition of 'natural' in terms of 'law' was important in another way. For if we did define 'natural' in terms of 'law', how would we then define 'law'? We couldn't use 'natural' in the definition (given the rules of directed definition). And Lewis thought—with some reason—that 'law' then simply couldn't be defined. So it would need to be taken as an undefined notion. And since lawhood is a sort of *modal* notion, it would be natural to view this as a metaphysics according to which reality is "ultimately" modal. Lewis's rejection of this sort of "modalism" is at the core of his metaphysics.

There are important questions to ask about the game of directed definition. What is the reason for saying that definitions can't go in circles? A very natural answer would appeal to the claim we discussed earlier: that in a (metaphysical) definition, the definiens is more natural than the definiendum. But that principle, as we saw, runs into trouble at USC. (Similar problems would arise with other principles of the form "the definiendum is less F than the definiens".) Is there any way of understanding why definitions can't go in circles that *would* pass muster at USC?[19]

Yes, there is. The key is this: what we are doing with definitions is often epistemic in nature, not metaphysical. And we can have different epistemic relations to the definiens and the definiendum. (Whether that's because the epistemic relations in question are really to *words*, or for some other reason, is not something I'm going to confront.)

For example, sometimes we play the game of directed definitions to address "discourse threat".[20] We find ourselves using some words, like 'law of nature'. We then confront a threat: maybe those words have some bad feature, like failing to be objective, or being massively vague, or even being meaningless. To confront this, we offer a definition of those words in terms of words that we take to be immune to the threat—words such that every sentence formed using them is objective, not massively vague, not meaningless, etc.

Definitions going in circles couldn't address discourse threat. If F is defined in terms of G, which is defined in terms of H, which is defined in terms of F, then if F may well be meaningless (or whatever), then all the others may well be meaningless as well. But using definitions to address discourse threat isn't undermined by coarse grain. Even if the proposition that such-and-such is a law just *is* the proposition that the best system includes such-and-such, the threat was to that proposition under the guise 'law of nature', not under the guise 'the best system includes…'.

For another example, sometimes we play the game of directed definitions to convince ourselves that a certain austere metaphysics is adequate. Lewis played it in support of his Humean supervenience: the view that every proposition supervenes on the "humean mosaic"—the spatiotemporal pattern of instantiation of perfectly natural local properties.[21] A natural worry about Humean supervenience is that true propositions of the form *it is a law that p* don't supervene on the Humean mosaic. This challenge can be answered by a definition of 'it is a law that' in terms of naturalness (provided naturalness itself supervenes on the mosaic!). And this method of answering the challenge isn't undermined

[19]In addition to what I discuss below, there's also accepting a somewhat less coarse-grained view, namely, Dorr's (2016) "only logical circles" principle.

[20]For a somewhat related discussion see Sider (2014).

[21]See Lewis (1986*b*, introduction); 1994.

by coarse grain for the same reason as we discussed in the previous paragraph: pre-challenge we perceive the proposition "it is a law that p" under one guise, and post-challenge we perceive that very same proposition under a different guise.

Before leaving the definition of naturalness in terms of law behind, it's worth noting that coarse grain raises a direct threat to the definition of natural properties and relations as those that figure in laws. What does "figuring in" mean? It seems to assume a structured conception of properties and relations, which is threatened by the Russell-Myhill paradox (see Fritz (2017) for an accessible presentation). We might avoid this by rephrasing the definition as this schema:

$$N(F) \text{ iff for some } G_1, \ldots, G_n\text{: } L(F, G_1, \ldots, G_n)$$

where N is 'is natural' (second-order predicate), the variables G_i can be of any type, and L is a sentential operator meaning "it is a law that". This would imply that every property F is natural, given coarse grain, since for any proposition p, $p = \big(p \wedge (F = F)\big)$.

4.3 Defining naturalness in terms of modal notions

Call a set of properties and relations *complete* iff all properties and relations supervene (globally, across all metaphysically possible worlds) on it. It's natural to assume (although see below) that the natural properties and relations are complete. (At least, assuming that the existence of properties isn't contingent. If it is, things get more complicated.) But then, one might think, one could define naturalness in terms of supervenience.

But what would be the definition? Not: "the set of natural properties is the complete set of properties and relations". For the set of *all* properties and relations is obviously complete.

What about "the set of natural properties and relations is the *minimally* complete set of properties and relations"—i.e. the complete set that contains no other complete set as a subset? The problem is that there won't be a unique minimally complete set, since if you start with one such set, you can mess around with it to generate another. Compare the set of truth functions: negation and conjunction is a minimal complete set of truth functions, but so is negation and disjunction (as are others). In mereology: you can start with parthood and define fusion and overlap; but you can also start with either of those two and define the others.

What about the idea that *every* member of every minimally complete set is perfectly natural? This will let in gruified versions of properties, since just as sets of properties or operations can have more than one "basis", as we discussed in the previous paragraph, they can also have a grueified basis. Suppose charge and mass are perfectly natural properties. Thus they are members of some minimal complete set. Now define the schmass of an object as its mass if it has unit negative charge and twice its mass otherwise, and replace mass in the set with schmass. The result will also be minimally complete; so schmass is perfectly natural.[22]

4.4 Melian definition

I'd like to mention one last possible definition of 'natural'. Suppose that physicalism is true, and in particular that the perfectly natural properties are in fact: charge, mass, and distance. (Set aside questions about the metaphysics of quantity.) Suppose we then offer this definition: p is a natural property iff p is either charge, mass, or distance. Would that be ok?

In Sider (2011, section 7.13) I called this view "Melianism". My main objection was that the predicates that figure in good explanations must express reasonably natural properties; 'is a property that is identical to either charge or mass or distance' does not express a reasonably natural property; therefore if Melianism is true, 'natural' cannot figure in good explanations.

I think it's a big deal whether Melianism is true. On one hand, the idea that in order to create the world (to use a common metaphor), God not only needed to utter a decree specifying objects' charges and masses (say), but also specifying that charge and mass are perfectly natural properties, is hard to swallow. On the other hand, the argument against Melianism feels compelling.

[22] Here is a more general argument. Suppose there is at least one minimal complete set, S. Now let u be any property. Taking our inspiration from Goodman (letting u play the role of "first observed before 3000"), for any p, define grue-p as $(p \wedge u) \vee (\sim p \wedge \sim u)$. Note that p is equivalent to $(u \wedge \text{grue-}p) \vee (\sim u \wedge \sim\text{grue-}p))$ (just as green is equivalent to "grue and first observed before 3000, or green and not first observed before 300"). So for any $p \in S$, if we remove p from S and then add u and also grue-p, the resulting set S' is still complete; u will therefore count as natural if S' is also minimal. Thus any u will count as natural if for some $p \in S$, the resulting set S' is minimal. I haven't tried yet to formulate a simpler version of this condition, but I don't think it's very restrictive.

5. Epistemology of naturalness

Armstrong (1978) claimed that it was an a posteriori matter which universals exist. Lewis didn't talk much about this issue, but he presumably thought the same thing about the question of which properties are natural. For each, the idea is presumably that we have reason to believe that predicates for well-confirmed scientific theories express natural properties.

Opponents of naturalness often make an epistemic objection: that if some "inflationary" metaphysics of naturalness is true (for instance if "primitivism" is true, or if naturalness is defined in terms of some other "metaphysical" notion, such as lawhood, which is itself not "reductively defined"), then we could not know which properties are natural. The idea, I think, is that given such a robust metaphysics of naturalness, it will always be possible that the predicates of a well-confirmed scientific theory do *not* express natural properties.

Objections like this to "inflationary" metaphysical views of all sorts are common. If antireductionism about personal identity were true, we would have no idea when persons are the same over time; if antireductionism about modality were true then we would have no idea what is possible; and so on.

There is a somewhat boring dialectic about the status of such objections. Fans of metaphysics point out how similar they are to arguments such as: "if the external world is truly external (as opposed to being composed of sense data, for example), then we would have no idea whether the external world exists". They go on to point out that if we're to avoid skepticism, knowledge (or justified belief) must be understood as being compatible with the possibility of certain sorts of error. Evidence that there exists an external table really does justify belief in the table even though the correct metaphysics of tables allows that the evidence is fallible; evidence that I lived in Philadelphia fifty years ago really does justify belief even if the correct metaphysics of personal identity implies that the evidence is fallible (or better: more fallible than it would be if that metaphysics were false); etc. And so: believers in an objective metaphysics of natural properties will say that the presence of predicates for natural properties in a well-confirmed scientific theory is evidence that those properties are natural, even though the evidence is fallible.

6. Objectivity, etc.

In a way, the most basic question about naturalness is whether it is objective. Is the distinction "in the world" or just "in us"?

We use the expressions 'green' and 'blue', and think of those as natural properties. But maybe we just "find them natural" because of some fact about us. Maybe there's nothing deeper going on beyond that English has a word for 'green' and not 'grue'. Or maybe it's a deeper fact about our culture, or history, or even biology; but still, it's something about us, not the world. (Nelson Goodman (1955*c*, 1978) held a view like this; Shamik Dasgupta (2018) is an important recent defender.)

Lewis, on the other hand, thought the distinction was objective (1983*b*, p. 347).

The question of objectivity is a central and important one for *any* metaphysical concept. Are the modal facts "out there", or merely in us? How about the facts about essence, or persistence over time, or causation? Or even existence (see Hirsch and Warren (2019); Thomasson (2015))?

To be sure, objectivity is a notoriously unclear notion. Many authors don't take up the question of objectivity for this reason (Lewis, for instance, says next to nothing about it). Nevertheless everyone effectively takes a stand on the question. Pick any metaphysician who talks about any topic; it's easy to figure out whether they think that the topic's questions are objective or not.

References

Armstrong, D. M. (1978). *A Theory of Universals*, volume 2 of *Universals and Scientific Realism*. Cambridge: Cambridge University Press.

— (1983). *What is a Law of Nature?* Cambridge: Cambridge University Press.

Beebee, Helen (2000). "The Non-Governing Conception of Laws of Nature." *Philosophy and Phenomenological Research* 61: 571–94.

Beebee, Helen and A. R. J. Fisher (eds.) (2020). *Philosophical Letters of David K. Lewis: Volume 1: Causation, Modality, Ontology*. Oxford, UK: Oxford University Press.

Bennett, Karen and Dean W. Zimmerman (2013). *Oxford Studies in Metaphysics*, volume 8. Ed. Karen Bennett and Dean W. Zimmerman. Oxford: Oxford University Press.

Dasgupta, Shamik (2018). "Realism and the Absence of Value." *Philosophical Review* 127: 279–322.

Dorr, Cian (2016). "To Be F is to Be G." *Philosophical Perspectives* 30: 39–134.

— (2019). "Natural Properties." Stanford Encyclopedia of Philosophy. Available at `https://plato.stanford.edu/entries/natural-properties/`.

Dorr, Cian and John Hawthorne (2013). "Naturalness." In Bennett and Zimmerman (2013), 3–77.

Fritz, Peter (2017). "How Fine-Grained is Reality?" *Filosofisk Supplement* 13(2): 52–7.

Gómez Sánchez, Verónica (2021). "Naturalness by Law." *Noûs* Forthcoming.

Goodman, Nelson (1955*a*). *Fact, Fiction, and Forecast*. Cambridge, MA: Harvard University Press.

— (1955*b*). "The New Riddle of Induction." In Goodman (1955*a*), 59–83.

— (1955*c*). "Prospects for a Theory of Projection." In Goodman (1955*a*), 84–131.

— (1978). *Ways of Worldmaking*. Indianapolis: Hackett.

Hall, Ned (2015). "Humean Reductionism About Laws of Nature." Manuscript. Available at `https://philpapers.org/archive/HALHRA.pdf`.

Hirsch, Eli and Jared Warren (2019). "Quantifier Variance." In *The Routledge Handbook of Philosophy of Relativism*, 349–357.

Lewis, David (1973). *Counterfactuals*. Oxford: Blackwell.

— (1983*a*). "Extrinsic Properties." *Philosophical Studies* 44: 197–200. Reprinted in Lewis 1999: 111–15.

— (1983*b*). "New Work for a Theory of Universals." *Australasian Journal of Philosophy* 61(4): 343–77. Reprinted in Lewis 1999: 8–55.

— (1984). "Putnam's Paradox." *Australasian Journal of Philosophy* 62: 221–36. Reprinted in Lewis 1999: 56–77.

— (1986*a*). *On the Plurality of Worlds*. Oxford: Blackwell.

— (1986*b*). *Philosophical Papers, Volume 2*. Oxford: Oxford University Press.

— (1991). *Parts of Classes*. Oxford: Blackwell.

— (1994). "Humean Supervenience Debugged." *Mind* 103: 473–90. Reprinted in Lewis 1999: 224–47.

— (1999). *Papers in Metaphysics and Epistemology*. Cambridge: Cambridge University Press.

Loewer, Barry (1996). "Humean Supervenience." *Philosophical Topics* 24: 101–27. Reprinted in J. Carroll (Ed.) (2004), *Readings on Laws of Nature* (Pittsburgh: University of Pittsburgh Press), pp. 176–206.

Marshall, Dan (2023). "Intrinsic vs. Extrinsic Properties." Stanford Encyclopedia of Philosophy. Available at `https://plato.stanford.edu/entries/intrinsic-extrinsic/`.

Putnam, Hilary (1978). *Meaning and the Moral Sciences*. Boston: Routledge and Kegan Paul.

— (1980). "Models and Reality." *Journal of Symbolic Logic* 45(3): 464–82. Reprinted in Putnam 1983: 1–25.

— (1981). *Reason, Truth and History*. Cambridge: Cambridge University Press.

— (1983). *Realism and Reason: Philosophical Papers, volume 3*. Cambridge: Cambridge University Press.

Schaffer, Jonathan (2004). "Two Conceptions of Sparse Properties." *Pacific Philosophical Quarterly* 85: 92–102.

Sider, Theodore (1993). *Naturalness, Intrinsicality and Duplication*. Ph.D. thesis, University of Massachusetts. Available at `http://tedsider.org/papers/diss.pdf`.

— (2007). "Parthood." *Philosophical Review* 116: 51–91.

— (2011). *Writing the Book of the World*. Oxford: Clarendon Press.

— (2013). "Against Parthood." In Bennett and Zimmerman (2013), 237–93.

— (2014). "Outscoping and Discourse Threat." *Inquiry* 57: 413–26.

— (2020*a*). "Crash Course on Higher-Order Logic." `http://tedsider.org/teaching/higher_order_20/higher_order_crash_course.pdf`.

— (2020*b*). *The Tools of Metaphysics and the Metaphysics of Science*. Oxford: Oxford University Press.

Thomasson, Amie L. (2015). *Ontology Made Easy*. New York: Oxford University Press.

Wallace, David (2019). "Who's Afraid of Coordinate Systems? An Essay on Representation of Spacetime Structure." *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* 67: 125–136.