

# Perfectly Natural Relative Naturalness

Billy Dunaway  
University of Oxford

## Abstract

This paper is about naturalness in Lewis’s sense. It motivates the need for a notion of relative naturalness, and defends the view that relative naturalness is among the perfectly natural properties by showing that such a view is consistent with a “Purity” constraint on perfectly natural properties as suggested by Sider (2012).

## 1 The problem: reference and relative naturalness

‘Naturalness’ in David Lewis’s sense is often said to play a number of important theoretical roles. Connections between supervenience, lawhood, and metaphysical possibility (among other things) seem to require the notion of a natural property in order to be adequately explicated.<sup>1</sup> The notion of naturalness that is needed here is one that makes a binary distinction between properties—those that are perfectly natural, and those that are not.

Other roles for naturalness require a *relative* notion, expressed by the relational predicate ‘is more natural than’. One of these is a plausible connection with reference-determination, which is roughly captured by the MAGNETISM thesis:

MAGNETISM The highly natural properties are easy to refer to.

MAGNETISM is attractive as a solution to familiar puzzles about indeterminacy in reference, including the “permutation argument” from Putnam (1981). A standard interpretation of our language on which ‘person’ refers to personhood, ‘banana’ refers to bananahood, etc. is one that “fits” our use of these terms fairly well: that is, it counts our utterances as by and large *true*. But there will be other “permuted” interpretations—that is, interpretations that assign referents which deviate from the standard interpretation in systematic ways—that score just as well as the standard interpretation in terms of fit.

According to MAGNETISM, reference is not however massively indeterminate between the referents assigned by standard and permuted interpretations: naturalness is, in addition to fit, a constraint on reference. On this line of thought,

---

<sup>1</sup>Cf. Lewis (1983). For a useful recent overview of the candidate roles for naturalness, see Dorr and Hawthorne (2013).

the standard interpretation does better at assigning highly natural referents to terms. The permuted interpretations are determinately incorrect interpretations on these grounds. Highly natural properties are then easy to refer to in the sense of MAGNETISM because reference is determined in part by the naturalness of candidate referents, in addition to considerations of fit.<sup>2</sup>

MAGNETISM is importantly a claim about *relative* naturalness. Personhood other referents assigned by the standard interpretation are more natural than the referents assigned by permuted interpretations, but they are not perfectly natural.

Lewis offered a way to understand relative naturalness in terms of the absolute notion of perfect naturalness. Each property has a *canonical definition*: a predicate containing perfectly natural terms only, plus some privileged connectives, that expresses the property in question. For Lewis, then, one property is more natural than another property iff its canonical definition is shorter. This is the DEFINITION thesis:

DEFINITION For any properties *A* and *B*, *A* is more natural than *B* iff the canonical definition for *A* is shorter than the canonical definition for *B*.<sup>3</sup>

Unfortunately combining DEFINITION with MAGNETISM has disastrous consequences. The motivations for introducing MAGNETISM include the thought that we can determinately refer to some properties of macro-level objects such as personhood. So these must be among the properties that are easy to refer to. But they will have extremely long canonical definitions, and so given DEFINITION these properties will not be very natural. MAGNETISM then won't explain why they are easy to refer to and Putnam-like indeterminacy puzzles will persist.<sup>4</sup>

## 2 Strategies for a solution

There are several broad naturalness-centric options for responding to this problem.

The first is to reject Lewis's characterisation of relative naturalness in DEFINITION, and to replace it with an alternative characterisation in the same spirit. Note

---

<sup>2</sup>This is a highly truncated version of Lewis's reply to Putnam in Lewis (1984).

<sup>3</sup>Here is Lewis:

Some few properties are *perfectly* natural. Others, even though they may be somewhat disjunctive or extrinsic, are at least somewhat natural in a derivative way, to the extent that they can be reached by not-too-complicated chains of definability from the perfectly natural properties. (Lewis (1986: 61), see also Lewis (1983: 347).)

<sup>4</sup>See Hawthorne (2006: 206).

that DEFINITION, by its own lights, makes relative naturalness out to be highly unnatural—it contains expressions like ‘term’ and ‘co-extensive’ which are not perfectly natural. Replacing them with their canonical definitions will result in an extremely long canonical definition for ‘is more natural than’, and hence make it out to be highly unnatural. Presumably other attempts at providing a definition in the same spirit will have a similar result. Call this the *Definitional Approach*.

The second is to give up on Lewis’s project of characterising relative naturalness in other perfectly natural terms. Instead of taking ‘mass’, ‘spin’, and the like as exhausting the catalogue of perfectly natural terms, we might instead take ‘is more natural than’ to be among the perfectly natural terms—relative naturalness is, on this view, perfectly natural. This approach eschews the need for a Lewis-style definition of relative naturalness. For those who are comfortable with the ideology, relative naturalness is “primitive”. Personhood, bananahood, etc. might be very natural, and MAGNETISM will be satisfactory if the relevant perfectly natural relation ranks them much higher than their counterparts in the fit-preserving permuted interpretations. Call this the *Perfectly Natural Approach*.

(A third approach is also a conceptual possibility, though I will mention it and then not discuss it further. This is the *Unprincipled Approach*. Take a ranking which, like the Perfectly Natural Approach, ranks personhood, bananahood, etc. as highly natural. Unlike the Definitional Approach, this is not because these properties score highly on some definitional characterisation of relative naturalness which makes it out to be highly unnatural. And unlike the perfectly natural approach, relative naturalness does not rank among the most natural properties. Instead, relative naturalness itself ranks as pretty natural.)

### 3 The Purity objection

Sider (2012) claims that the correct approach here is a version of the Definitional Approach, as he has an argument that purports to show that the Perfect Naturalness Approach is not an option.<sup>5</sup> (He does not consider the Unprincipled Approach.) The core of the argument centres around his “Purity” constraint, which he motivates with the following picture:

When God was creating the world, she was not required to think in terms of nonfundamental notions like city, smile, or candy.<sup>6</sup>

On this picture, we imagine God settling the facts about cities, smiles, etc.; suppose this is done by a tokening of a mental sentence in a specific box in the Divine

---

<sup>5</sup>Sider doesn’t actually provide an alternative to DEFINITION; instead, he claims only that *some* definitional approach must be workable but that we should not be put off by the fact that it is difficult to provide one (Sider 2012: xx).

<sup>6</sup>Sider (2012: 106)

Mind. (This is somewhat analogous to the familiar “belief box” but is a unique psychological feature of the Divine Mind, as the facts expressed any sentence tokened in it automatically pop into being.) To successfully create, God needs to token in this box sentences that specify the distribution of all of the perfectly natural properties. But the constraint Sider proposes requires that all of the tokenings be couched in perfectly fundamental terms; the Divine Mind cannot be required to traffic in nonfundamental notions when settling how the perfectly natural properties are distributed.

This of course is highly picturesque. I don’t wish to defend it here. Instead, I will simply assume that something like the intuitive idea behind Sider’s picture is correct, and ask what its consequences for the Perfectly Natural Approach are.

One consequence of this picture is that something like STRONG PURITY is true:

STRONG PURITY For every perfectly natural property  $P$ , every fact involving  $P$  is perfectly natural.<sup>7</sup>

Here, a “perfectly natural fact” is one which can be expressed by a sentence where all the predicates in the sentence stand for perfectly natural properties; if every sentence which expresses the fact uses not-perfectly natural predicates, then the fact is not perfectly natural. (I will leave the questions of what facts are, and of what it is for a property to be “involved” in a fact unanalysed. Also one might extend ‘natural’ to expressions beyond those which express properties; I will not address this question either.)

STRONG PURITY is not uncontroversial. It requires that every one of the facts involving perfectly natural properties be perfectly natural. But this is stronger than what the motivating picture requires: God would be able to “fix” the distribution of mass by tokening only names perfectly natural properties even if the fact that John thinks about mass isn’t perfectly natural. One might try to formulate weaker Purity-style theses which make place restrictions exactly which facts involving perfectly natural properties need to be perfectly natural. Instead of trying to do this here, I will simply assume an informal notion of the “distribution-fixing” facts for a property, and read STRONG PURITY as requiring that, if relative naturalness is perfectly natural, then all of the *distribution-fixing* facts about perfect naturalness must be perfectly natural. Since my aim here is to investigate the consequences for the Perfectly Natural Approach, I will simply spot the Purity-

---

<sup>7</sup>Compare Sider’s original “Purity” constraint on Sider (2012: 106), which holds that “fundamental facts involve fundamental notions only.” Sider does not use the ideology of ‘naturalness’, but his motivations have consequences for the naturalness-centric framework—though it is not always immediately obvious how best to formulate them. I will work with what I take to be plausible reconstructions below, though a more thorough investigation of how the translations should go would be of independent interest.

lover this informal restriction, even if there are substantial questions about how it should be understood.

STRONG PURITY, even once it is appropriately restricted, would appear to have disastrous consequences for the Perfectly Natural Approach. This is because the additional premise which I will call LOGIC OF RELATIVE NATURALNESS is quite plausible:

LOGIC OF RELATIVE NATURALNESS For any properties  $A$  and  $B$ , if  $A$  is more natural than  $B$  then  $B$  is not perfectly natural.

The reasoning behind LOGIC OF RELATIVE NATURALNESS is straightforward: if  $A$  is more natural than  $B$ , then there is something more natural than  $B$  and hence  $B$  is less-than-perfectly natural.

Take any fact about relative naturalness expressed by a sentence of the form  $\lceil A$  is more natural than  $B \rceil$ . This would appear to be a distribution-fixing fact about relative naturalness in the informal sense from above: if relative naturalness is perfectly natural, then God wouldn't be able to settle all the relative naturalness facts without tokening in the Divine Mind a sentence that expresses  $A$ 's being more natural than  $B$ . But the following argument against the Perfectly Natural Approach then presents itself. By LOGIC OF RELATIVE NATURALNESS,  $B$  is not perfectly natural. Then, there is some distribution-fixing fact—namely the one expressed by  $\lceil A$  is more natural than  $B \rceil$  which is not perfectly natural. Hence by STRONG PURITY relative naturalness is not perfectly natural.<sup>8</sup>

#### 4 Horizontal and vertical relative naturalness

Even when we spot the Purity-lover an informal restriction to the distribution-fixing facts, the argument from STRONG PURITY works only if we assume that any property which is not perfectly natural is not involved in perfectly natural facts. That is:

PROPERTY-FACT CONNECTION If  $P$  is not a perfectly natural property, then any fact involving  $P$  is not perfectly natural.

---

<sup>8</sup>Cf. Sider's conclusion:

On any conception of the nature of the connection, connecting truths—true statements expressing the distinctive connection between fundamental and nonfundamental matters—are going to [...] involve nonfundamental notions. [...] So on any conception of the nature of the connection, the connecting truths are going to have to be nonfundamental. (Sider (2012: 110))

Recall moreover that a perfectly natural fact is one which can be expressed by a sentence where all the predicates in the sentence stand for perfectly natural properties. *Prima facie* there is a distinction here: whether a property is perfectly natural is a different question from whether it can be expressed using perfectly natural terms only.

As it turns out, PROPERTY-FACT CONNECTION is true on some ways of understanding relative naturalness, but false on others.

We can distinguish between two notions of relative naturalness which differ in structurally important ways. One is a “vertical” notion of relative naturalness, on which some of the relative naturalness facts do involve necessarily co-extensive properties. For instance, on this approach water and H<sub>2</sub>O might be natural to different degrees even though they are necessarily co-extensive. That is:

VERTICAL RELATIVE NATURALNESS (VRN) For any property *A*: if *A* is not perfectly natural, then there is some property *B* that is necessarily co-extensive with *A*, and *A* is more natural than *B*.

Even if two properties share an intension, on a VRN view, there might nonetheless be a difference in naturalness between them. Here is a simple example. Let a “Marge” be something that either has mass or has charge. On a vertical naturalness view, Margehood is not perfectly natural and so there is a necessarily co-extensive perfectly natural property—the disjunctive property *having mass or charge*—that is more natural than it.

A VRN view requires a hyperintensional view of properties, as properties which share an intension but differ in naturalness must be distinct properties. Thus Margehood and *having mass or charge* are distinct.

VRN views are incompatible with STRONG PURITY, since the defining structural feature of vertical naturalness relations guarantees that PROPERTY-FACT CONNECTION is true. Return to our simple example. If a fact involving Margehood is to be a perfectly natural fact, then it must be possible to express it using only predicates that stand for perfectly natural properties. But to do this, we must use the expression ‘having mass or charge’. And since the VRN view requires a hyperintensional distinction among these properties, the expression ‘having mass or charge’ will *not* serve to express a fact involving Margehood—instead it will express the distinct property *having mass or charge*. Thus facts about the relative naturalness of Margehood cannot be expressed using perfectly natural expressions only. Since similar reasoning applies for other not-perfectly-natural properties, PROPERTY-FACT CONNECTION will be true on a vertical view of relative naturalness.

The other option is a “horizontal” view of relative naturalness, where distinctions in degree of naturalness only occur among properties that differ in intension. That is:

HORIZONTAL RELATIVE NATURALNESS (HRN) For any properties *A, B*: if *A* is more natural than *B*, then *A* and *B* are not necessarily co-extensive.

Lewis's own view took relative naturalness to be a horizontal relation. Distinctions in naturalness can only obtain between properties that differ in intension, on his view, because differences in length of canonical definition will only obtain between properties that differ in intension. One consequence of the structural differences between VRN and HRN views is that, while the former require hyperintensional distinctions among properties, the latter is compatible with a simple view which, like Lewis, identifies properties with their intensions.

Margehood is not perfectly natural, but the relative naturalness relations it stands in are have different structural features than on the VRN view. What is more natural than it is not some necessarily co-extensive property but rather (for instance) the non-co-extensive property of mass; the horizontal theorist treats sentences like 'mass is more natural than Margehood' as expressing the paradigmatic relative naturalness facts.

Because of this, PROPERTY-FACT CONNECTION is false on some HRN views. The HRN view can maintain that Margehood, which not-perfectly-natural, is expressed by its canonical definition, 'having mass or charge' (the property expressed by this predicate is necessarily co-extensive with Margehood, and there is no barrier to the HRN view holding that they are therefore identical). All of the predicates in the canonical definition are perfectly natural terms. Then, for any fact involving Margehood, there is a way of expressing the fact that uses the canonical definition of Margehood to express the same fact. So though Margehood is not perfectly natural, the facts that involve it are, as they can be expressed using the canonical definition of Margehood. PROPERTY-FACT CONNECTION is false given a horizontal structure to relative naturalness.

More generally, relative naturalness will satisfy STRONG PURITY on such a view. Any distribution-fixing fact expressed by a sentence of the form  $\lceil A \text{ is more natural than } B \rceil$  will involve a not-perfectly-natural fact, namely  $B$ . This is what LOGIC OF RELATIVE NAURALNESS tells us. But it doesn't follow that the fact itself is not perfectly natural, since the same fact can also be expressed using the canonical definitions of  $A$  and  $B$ . And the only predicates used by these canonical definitions are predicates standing for perfectly natural properties. The relative naturalness facts will be perfectly natural.

The upshot is that the proponent of the Perfectly Natural Approach still has a viable approach to MAGNETISM. She can hold that the relative naturalness relation is perfectly natural, and, like Lewis, hold that it is a horizontal relation. The canonical definitions of a property can be used to express the relative naturalness facts without violating STRONG PURITY. The only point of departure is that she, unlike Lewis, holds that the relative naturalness relation is not hostage to the features of canonical definitions; it is instead a perfectly natural relation.<sup>9</sup>

---

<sup>9</sup>Thanks to John Hawthorne and David Manley for helpful discussion of previous versions of this paper.

## References

- Cian Dorr and John Hawthorne. Naturalness. In Karen Bennett and Dean Zimmerman, editors, *Oxford Studies in Metaphysics vol. 8*. Oxford University Press, 2013.
- John Hawthorne. Epistemicism and Semantic Plasticity. In *Metaphysical Essays*, pages 185–210. Oxford University Press, 2006.
- David Lewis. New Work for a Theory of Universals. *Australasian Journal of Philosophy*, 61(4):343–377, 1983.
- David Lewis. Putnam’s Paradox. *Australasian Journal of Philosophy*, 62(3):221–236, 1984.
- David Lewis. *On the Plurality of Worlds*. Basil Blackwell, 1986.
- Hilary Putnam. *Reason, Truth and History*. Cambridge University Press, 1981.
- Theodore Sider. *Writing the Book of the World*. Oxford University Press, 2012.