Humility and Constraints on O-Language[*]

Stephan Leuenberger

According to scientific realism, it is one the ambitions of science to discover the fundamental properties. If the argument in David Lewis's "Ramseyan Humility" is sound, however, that ambition must remain unfulfilled: we are irremediably ignorant about what these properties are.

That surprising conclusion has already given rise to a sizeable literature, which includes Langton [2004], Schaffer [2005], Whittle [2006] and Ney [2007]. These responses discuss Lewis's ignorance thesis in the light of more general issues in the theory of knowledge and the metaphysics of properties. In contrast, some of the semantic assumptions that Lewis relies on have so far received little scrutiny. In this paper, I aim to make them explicit, and argue that they are false.

"Ramseyan Humility" can be read as continuing a debate about what the content of scientific theories is. It was once widely held that the vocabulary of the language of science divides into theoretical and non-theoretical terms. On this view, the content of the theory is, roughly, given by its Ramsey sentence, which does not contain any theoretical terms.[1] For theoretical terms are taken to be definable by non-theoretical ones, and definitions, being merely abbreviations, do not increase the expressive power of a language.

Is such a conception of scientific language and theories compatible with scientific realism? For a scientific realist, it ought to be a substantive claim that a given scientific theory is true, and there ought to be a fact of the matter about what entities are referred to by theoretical terms. The so-called model-theoretic arguments, championed by Hilary Putnam (e.g. in Putnam [1980] and Putnam [1981]), threaten scientific realism thus construed. They raise the double specter of triviality and indeterminacy. Triviality, because it seems that every consistent theory comes out true. Indeterminacy, because a true theory does not seem to uniquely determine the referents of subsentential expressions. In Lewis [1983a] and Lewis [1984], Lewis is concerned with the threat of triviality. He argues that

if only relatively natural, non-disjunctive properties are eligible to be referents, then there is no danger that every consistent theory comes out true. In Lewis [forthcoming], he in effect concedes that this move does not help to foil the threat of indeterminacy. There would not be a unique referent even if only perfectly natural properties were eligible. Moreover, he proclaimed not to be worried by such indeterminacy.

In this paper, I do not try to settle whether we ought to endorse Lewis's conclusion, and whether we ought to be worried if it were true. Such questions would lead us deeply into the theory of knowledge and the metaphysics of properties. My project here is limited to showing that the particular argument for the indeterminacy that Lewis offers in "Ramseyan Humility" does not work.

The plan of the paper is as follows. In section 1, I try to pin down the thesis of Humility that Lewis is arguing for. Section 2 presents a reconstruction of the argument in deductive form, and clarifies some terminology. Section 3 attacks what I call "Structuralism," a premise about the expressive power of our language minus its theoretical terms. Finally, section 4 considers and criticizes strategies to rescue Lewis's argument.

1   The Thesis of Humility

The conclusion of Lewis's argument is named after a virtue: it is the thesis of "Humility."[2] But Lewis does not give us a canonical formulation of that thesis. The first problem in reconstructing the argument is to determine what exactly its conclusion is. What sort of knowledge concerning fundamental properties and their roles are we supposed to lack? According to Lewis, we cannot know "the identity of the properties ..." (p. 12); the "true contingent proposition about which of the possible realizations is actual" (p. 5/6); and the answer to "the question: which property occupies the role?" (p. 13).

Examining Lewis's arguments provides us with a better clue to what the thesis of Humility is than these formulations do. He offers two different, though closely related arguments: the Replacement Argument and the Permutation Argument. The former concludes that we cannot know which ones among the possible fundamental properties and relations are actually instantiated. (I assume that for a fundamental property to exist in a world and for it to be instantiated in a world are the same thing.) The latter concludes that we cannot know which ones among the actual fundamental properties play which roles.

I will only analyse the Replacement Argument in detail, but my diagnosis carries over to the Permutation Argument, *mutatis mutandis*. The latter deploys premises that are more likely to be acceptable to certain actualist philosophers.[3] However, I set actualist scruples aside in this paper and grant Lewis that we can freely quantify over non-actual entities. Since the permutation argument does

not offer advantages other than greater acceptability to actualists, it does not require separate discussion here.

Lewis allows that Humility may apply to some fundamental properties and relations but not to others. Perhaps he thinks that it applies to mass and charge, but not to spatiotemporal distance. But if it applies to F, then it will likewise apply to every G in the same category as F, if Lewis's arguments are sound. He introduces the notion of a category, to which Humility is relativized, as follows:

> I speak of 'fundamental properties' for short, but they fall into several categories. There are all-or-nothing monadic properties. There are all-or-nothing n-adic relations, at least for smallish n. There are properties that admit of degree, that is, magnitudes; more generally, there are scalar-valued, vector-valued, tensor-valued, ... magnitudes. There are relational magnitudes. Maybe my list is too long; maybe the magnitudes could somehow be reduced to all-or-nothing properties and relations, but that is a question I shall not take up here. (Lewis [forthcoming, p. 3])

It emerges from Lewis's discussion that he takes it to be non-contingent what category a given fundamental property belongs to. For example, if mass is actually a scalar-valued fundamental property, it could not be vector-valued, or even a relation, in some other world.

For present purposes, let Humility about category C be the thesis that we do not know which members of C are actual. On standard views, that S knows which things are ϕ implies that S knows that x is ϕ, if x is one of the things which are ϕ. I am assuming here that this is correct.[4] Therefore, Humility about C follows from the claim that for any fundamental property F that belongs to C, we cannot know that F exists.

2 The Replacement Argument for Humility

Here is a simplified version of the Replacement Argument: Consider a world w that is just like the actual world @, except that whenever a particular value of mass is instantiated in a region of actual spacetime, the corresponding value of some non-actual fundamental property, schmass, is instantiated in the corresponding region of w. Then we cannot tell whether we are in @ or in w, the argument goes, and we cannot know that we are in a world where mass is instantiated. The same point can be made about other putatively fundamental properties: we can consider a world where charge is replaced by schmarge, and where spin is replaced by schpin. The Permutation Argument considers a world where mass is instantiated in all the locations where charge is in the actual spacetime, and *vice versa*, and otherwise proceeds much like the Replacement Argument.

Lewis's actual arguments are more complex than the simplified versions just given. I present a reconstruction of his Replacement Argument in deductive form.

While I deviate to some extent from his terminology, I take the reconstruction to be faithful.[5] However, I do not attempt to defend this exegetical claim until section 3 (3.1).

In my reconstruction, there are three key premises: First, an *expressibility premise*: knowledge must be expressible in the language O, to be introduced below; or more weakly, a knowable proposition must be logically entailed by a proposition expressible in O. Second, a *structuralist premise*: O-language cannot discriminate between worlds with the same structure. Third, a *combinatorialist premise*: fundamental properties are recombinable in such a way that there are distinct possible worlds that have the same structure.

These three claims appear as premises 1), 2), and 3), respectively, in the following valid argument:

1) If A is knowable, then it is entailed by a true proposition that is expressible in O.[6]

2) If A is expressible in O, then it supervenes on fundamental structure.

3) If not every member of category C is actual, and A is a true proposition that entails that $F \in C$ exists, then A does not supervene on fundamental structure.

4) Not every member of category C is actual.

5) If A is true and entails that $F \in C$ exists, A does not supervene on fundamental structure. (from 3) and 4))

6) If A is true and entails that $F \in C$ exists, A is not expressible in O. (from 2) and 5))

7) If A entails that $F \in C$ exists, A is not knowable. (from 1) and 6))

The conclusion 7) entails that we cannot know that F exists, for any $F \in C$. As noted above, this claim in turn entails Humility about C.

The intermediate conclusion 6) is a version of what is sometimes called "quidditism": the claim that two properties in two different worlds can play exactly the same role. On one way of cashing out talk of "roles", two properties share a role if O cannot distinguish them. According to 6), whatever can be truly said in O about F in world w is also true about some F' in a world w' where F does not exist.[7]

Some of the premises require comment.

Since Humility is relativized to categories, the argument is schematic, and for all we know, premise 4) may be true for some choices of a category C but not for others. Perhaps the spatial and the temporal distance relations are the only

members of the category of scalar-valued two-place fundamental relations. If so, we can know that these relations exist, for all the replacement argument says. But since we have set actualist scruples aside, we can grant Lewis that premise 4) is true for at least some interesting categories C.

This leaves us with the three substantive premises introduced above: the expressibility premise 1), the structuralist premise 2), and the combinatorialist premise 3).

Premise 3) may not initially look like a statement of combinatorialism. However, it is a consequence of a general combinatorial principle for fundamental properties and relations. Suppose the antecedent of 3) holds. Then there is a G, of the same category as F, that is not instantiated in the actual world. Combinatorialism guarantees that there is a world w in which the fundamental properties and relations are distributed just like in the actual world, except that F is throughout "replaced" by G. Then the proposition that F exists is true in the actual world, but false in the fundamentally isomorphic w, and consequently does not supervene on fundamental structure. Any proposition that is true and supervenes on fundamental structure is true in w, and hence does not entail that F exists. Thus the consequent of 3) holds as well.

Premises 1) and 2) require clarification in two respects. What is it for a proposition to supervene on fundamental structure? And crucially, what is the language O? In the rest of this section, I answer the first of these questions, and discuss the second in a preliminary way. The content of 1) and 2) will be further clarified in section 3.

What is it for a proposition to supervene on fundamental structure? The notion of supervenience relevant for premise 2) is *global* supervenience. A proposition A globally supervenes on fundamental structure if whenever worlds w and w' have the same fundamental structure, then the truth-value of A is the same in w and in w'. What it is to have the same fundamental structure can be cashed out using the notion of an isomorphism. Let F_w and $F_{w'}$ be the classes of fundamental properties instantiated in w and w', respectively, and let D_w and $D_{w'}$ be the domains of individuals of these worlds. Then a one-one function f from $D_w \cup F_w$ onto $D_{w'} \cup F_{w'}$ is a *fundamental isomorphism* from w to w' if it preserves instantiation structure in both directions, i.e. if the following holds for every $x \in D_w$ and $X \in F_w$: x instantiates X in w iff $f(x)$ instantiates $f(X)$ in w'.[8] The existence of a fundamental isomorphism is a necessary and sufficient condition for worlds to have the same fundamental structure.

Sameness of fundamental structure has been explained in terms of fundamental properties. While we cannot define what it is for a property to be fundamental, we can clarify it to some extent. It would be best to start with examples, but unfortunately, we cannot identify any clear cases. Even supposing that Humility is false, and that it is thus knowable what the actual fundamental properties are,

it is unlikely to be known at the present stage of inquiry. Lewis conjectures that mass and charge might be fundamental, and would thus be referred to even in a final and complete scientific theory of our world. Arguably, some spatiotemporal relation is fundamental too, but it is not clear exactly which relation. While there are no clear paradigm cases, there are clear foils. Most properties we ordinarily talk about, such as being red or being metallic, are non-fundamental. From the reduction of thermodynamics to statistical mechanics, we can conclude that degrees of temperature are not fundamental either. Likewise, properties that are intuitively negative or disjunctive or relational fail to be fundamental. As Lewis puts it, fundamental properties and relations are "perfectly natural." Moreover, part of what it is to be fundamental is to belong to a minimal global supervenience base for all classes of properties.[9]

The notion of supervenience on fundamental structure is relatively unproblematic in the context of this argument. More issues will be raised by the question what the language O, referred to in premises 1) and 2), is supposed to be. As I noted in the introduction, Lewis is appealing to a distinction between two types of terms, theoretical and non-theoretical. The language in which the final and complete theory of the world is stated contains terms of both sorts, and O is the fragment of that language which does not include the theoretical terms.

What are the non-theoretical terms? Traditionally, they have been taken to be observational terms. But in Lewis [1970], Lewis proposed that we take 'O' to mean 'old' rather than 'observational'. The terms in O are those that are understood pre-theoretically.

Given its prominent role in the argument, it is surprising how little Lewis says about what features O is meant to have. Maybe he thought that his argument is schematic in O, and would be sound for any choice of O that meets the constraints he explicitly lays down. But as will be shown in detail in section 3, it is not sound for every such choice, but depends crucially on the expressive power of O.

What constraints does Lewis impose? He assumes that fundamental properties are not named in O "except as occupants of roles." The nature of this constraint will be discussed in section 3. Further:

> I'll assume ... that O-language does suffice to express all possible observations, whatever else it may also be able to do. I'll assume that O-language is interpreted—never mind how." (Lewis [forthcoming, p.3])

Lewis is adamant that he does not impose certain other constraints on O that one might expect:

> 'O' stands for 'old'; it is the language that is available to us without the benefit of the term-introducing theory T. 'O' does not stand

for 'observation'. O-language is not meant to be a 'pure observation language', and indeed I doubt that there could be any such thing...

I do not assume it to be first-order or extensional or finitary or free of indexicality. Nor do I assume it to be unmetaphysical, suited only to talk of everyday matters. (pp. 3-4)

After reconstructing Lewis's replacement argument and clarifying some of its premises, I will critically examine it in the rest of the paper. I will argue that 2), the structuralist premise, is not explicitly acknowledged by Lewis, let alone adequately defended. This premise guarantees that a world where the actual fundamental properties are replaced has the same complete description in the language O. I will argue that this premise is not just unacknowledged and unsupported, but also highly implausible.

3   The Structuralist Premise

The achilles' heel of the Replacement Argument, I claim, is premise 2). I will refer to it as the thesis of "Structuralism," despite the fact that this term has other uses in different philosophical debates:

Structuralism If a proposition is expressible in O, then it supervenes on fundamental structure.

Informally, Structuralism can be taken as the doctrine that the language O does not capture anything over and above the structure of the world.

Does Structuralism have any *prima facie* plausibility, or if not, is it supported by theoretical considerations?

Before arguing for a negative answer to this question, I need to make a case that the Replacement Argument indeed relies on Structuralism, even though Lewis does not explicitly acknowledge 2) or anything equivalent to it as a premise.

3.1   Lewis's implicit appeal to Structuralism

To answer the worry that I am reconstructing the argument uncharitably, I will identify the transition in which 2) must work as a background assumption. While I claim that Structuralism is indispensable for Lewis's argument, I do not claim, of course, that any argument for Humility would have to rely on it.

Lewis deploys the notion of an n-tuple of properties realizing a final and complete theory T of our world. T itself is not formulated in O: it contains property names $P_1, ..., P_n$ that are not part of O's vocabulary. If we substitute these property names with variables, we obtain an open sentence T^O which belongs to O. T^O is the matrix of the Ramsey sentence T^R of T; i.e. T^R is

existentially quantified with respect to all variables free in T^O. To say that an n-tuple realizes T in a world w is to say that it satisfies the open sentence T^O in w.

In the following passage from Lewis's statement of the Replacement Argument, he is in effect trying to establish the intermediate conclusion 6) of my reconstruction: that no truth expressible in O entails that F exists, since even a complete theory formulated in O cannot distinguish between a world where F exists and a world where it has been replaced.

> We start with the unique actual realization of T; all fundamental properties except idlers and aliens are members of it. If we replace those properties by others, we get a possible realization by combinatorialism. (p. 10)[10]

If $\sigma = \langle F_1, ..., F_n \rangle$ is an n-tuple of fundamental properties and relations, a *replacement* $\sigma' = \langle F'_1, ..., F'_n \rangle$ of σ is an n-tuple such that F_i and F'_i belong to the same category (for $1 \leq i \leq n$). If σ consists of all and only the fundamental properties of w, $\sigma'w$ is the world fundamentally isomorphic to w where the members of σ have been replaced by the members of σ'. Combinatorialism guarantees that there is such a world; and the fact that the fundamental properties and relations form a global supervenience base for everything guarantees that it is unique.[11]

In the above quote, Lewis is talking about arbitrary replacements, not about one particular replacement. Hence he is committed to the claim that the actual world satisfies the following condition, where T_w is the true and complete theory of world w:

2′) If σ realizes T_w in w, then every replacement σ' of σ realizes T_w in $\sigma'w$.

Note that the consequent asserts that σ' realizes T_w in $\sigma'w$, not merely the triviality that it realizes $T_{\sigma'w}$ in $\sigma'w$.

There might be a different argument for distinct possible realizations of T_w that relies only on the corresponding existential claim: that there exists at least one replacment σ' that realizes T_w in $\sigma'w$. However, such an argument would be different from Lewis's; I will discuss its prospects in section 4.2.

Is 2′) supposed to reveal a contingent feature of the actual world, or does it state a condition that all worlds fulfill? In other words, is 2′) to be asserted only when the actual world is assigned to the variable 'w', or is it to be understood as prefixed by a tacit universal quantifier 'for all worlds w'?

If one accepted 2′) as true of the actual but not of all worlds, then clearly some argument would be needed. Otherwise the assertion of 2′) as true of the actual world is *ad hoc*. Why should the fundamental structure of the actual world be enough to fix all O-facts, if the fundamental structure of some other

worlds does not? What is special about the fundamental structure of our world that precludes O-variation?

Lewis does not engage in any such special pleading on behalf of the actual world. Indeed, it would be odd for him to do so, given his insistence that we have no reason to think that the our world is metaphysically distinguished. He complains against the view that in ours, but not in every world, all fundamental properties are instantiated: it posits that "our world has a special distinction that some other worlds lack. Why should the world we happen to live in be special in this way? Very likely it isn't" (p. 10).

I will thus assume that Lewis accepts 2′) as a claim about the actual world because he accepts 2′) as a claim about all worlds. While this assumption simplifies the discussion in 3.2, 3.3, and 3.4, it is dispensable, as I argue in 3.5.

Structuralism is a consequence of universally quantified 2′). Suppose that w and w' are fundamentally isomorphic worlds. Then the realization of the true and complete theory $T_{w'}$ of w' is a replacement σ' of the realization σ of the true and complete theory T_w of w. The sequence σ realizes T_w in w, and by 2′), σ' realizes T_w in $\sigma'w$. Hence T_w^R is true in both w and $\sigma'w$. Since T_w is a complete theory, T_w^R entails every O-sentence that is true in w.[12] Hence exactly the same O-expressible propositions are true in w and $\sigma'w$. Since w and w' were chosen arbitrarily, this shows that O-expressible propositions cannot differ among fundamentally isomorphic worlds; that is, they supervene on fundamental structure.[13]

After arguing that Lewis's Replacement Argument does indeed rely on Structuralism, it remains to examine the status of the latter.

3.2 Structuralism is implausible

Clearly, Structuralism is a substantive claim about O: there certainly are possible languages that can express truths that fail to supervene on fundamental structure. Suppose that some constant P denotes a fundamental property F that is instantiated in @. Then the sentence '$\exists x \, P(x)$' is true in @, but not in a fundamentally isomorphic world where F has been replaced.[14] Hence not every truth expressible in that language supervenes on fundamental structure.

Lewis rules out that some constant of O denotes a fundamental property, and has thus nothing to fear from counterexamples of that sort. In 3.4, I discuss whether his explicit assumptions about O entail Structuralism. Before, I argue that Structuralism is false if O is anything like our language.

If there are examples of O-expressible differences between structually isomorphic possible worlds, the claim that every O-expressible proposition supervenes on fundamental structure is false. The examples that I will suggest involve exotic examples of fundamental properties and relations, which most of us think

are not instantiated in actuality. This is legitimate in the dialectical context, since the Replacement Argument explicitly asks us to consider alien fundamental properties.

However, I can see no reason to think that Structuralism is true when we restrict our quantifiers in the supervenience claim to worlds that are nearby in modal space, say to worlds in which no alien fundamental properties are instantiated, or to worlds of a similar complexity as ours. The reason I focus on small and exotic worlds is that they are much easier to describe. I offer a variety of examples, each one of which is contentious, and may be rejected by philosophers with certain theoretical commitments. Nonetheless, pointing to different apparent counterexamples to Structuralism should make it plausible that at least one of them is a real counterexample.

Example 1: Physical vs. non-physical.
Let w_p be a world in which there is just one point-sized particle that exists for a given period of time, and throughout instantiates the fundamental property of having a mass of 1 milligram. Let w_i be a world in which there is one point-sized immaterial thing, which exists for the same amount of time, and instantiates exactly one fundamental non-physical property. The worlds w_p and w_i are fundamentally isomorphic.

Sentences such as "there is a material thing" and "some physical property is instantiated" express propositions that are true in w_p and false in w_i. If either of the terms 'material' or 'physical' belongs to the language O, or has a synonym in O, then these propositions can be expressed in O, and Structuralism is false.

Example 2: Phenomenal differences.
Take your favorite paradigms of a phenomenal experience, say being in pain and having an experience as of seeing red. Suppose that in some worlds, things have such experiences not in virtue of having a complex arrangement of various properties and relations, but in virtue of instantiating fundamental properties F and G, respectively. (On some views, this is consistent with physicalism being true in our world.) Let w_F and w_G be worlds in which there is just one thing instantiating one fundamental property, F and G respectively. Then w_F and w_G are fundamentally isomorphic.

The sentence "something is in pain" expresses a proposition that is true in w_F and false in w_G. If 'in pain' or a synonym is in O, then that proposition can be expressed in O, and Structuralism is false.[15]

Example 3: Edenic colour differences.
Let F and G now be what David Chalmers calls "edenic" colour properties, say perfect red and perfect blue. These are non-actual fundamental properties, not role properties that need a realizer. As in Example 2, there is a pair of fundamentally isomorphic worlds that differ only as to whether F or G is instantiated.[16]

The sentence "something is red" expresses a proposition true in one of these worlds but false in the other. If 'red' or a synonym is in O, then that proposition can be expressed in O, and Structuralism is false.

Examples 1 to 3 involve pairs of fundamentally isomorphic worlds that differ with respect to which fundamental monadic properties are instantiated. Plausibly, such pairs might instead differ with respect to which fundamental relations are instantiated. On Lewis's view, there are no actual fundamental relations except spatiotemporal ones. But perhaps there are other fundamental relations in other possible worlds. To use an example due to Hawthorne [2001], there might be a perfectly natural relation of attending, such that whether one entity attends to another is not constrained by the intrinsic properties of the relata or by the spatiotemporal relations between them. There might be a possible dyadic fundamental relation of psychic connection that comes in degrees, and that in some worlds is distributed isomorphically to distance relations in our world. If we accept that there could be such fundamental relations, we can construct further *prima facie* counterexamples to Structuralism.

Example 4: Non-spatiotemporal fundamental relations
Let w_D be a world where there are two things, each instantiating one fundamental monadic property, at a small distance from each other. Let w_T be a world where there are two non-spatial things, each instantiating one fundamental monadic property, in some fundamental intentional relation in the same category as distance.[17] Then w_D and w_T are fundamentally isomorphic.

Then there are sentences, such as "two things are apart from each other, but close," that express propositions that are true in w_D and false in w_T. If terms like 'apart' and 'close', or synonyms, are in O, then such propositions can be expressed in O, and Structuralism is false.[18]

It is easy to construct further examples, some more plausible than others. Perhaps we can even describe a pair of fundamentally isomorphic worlds that have a different causal structure.[19] If there are causal locutions in O, then this will give

rise to further O-expressible propositions that do not supervene on fundamental structure.

Of course, there are several ways to resist the conclusion that Structuralism is false. First, for each example it could be objected that I have failed to describe a pair of possible worlds. Secondly, my semantic ascent could be challenged: the sentences of English that I mention do not express propositions with different truth-values in the two worlds. Thirdly, it could be denied that the terms I mention belong to O.

It turns out that the second and third of these objections are of no help to the proponent of the Replacement Argument. Explaining why will require a closer examination at how the argument works, which I provide in 3.3.

The first objection is hard to refute conclusively. Roughly, my claim that the worlds I described are possible is based on their conceivability and a presumption that what is conceivable is also possible. Justifying the possibility claims in detail would involve a general discussion of modal epistemology that is beyond the scope of this paper. But I think it is fair to say that Lewis, of all philosophers, was willing to accept that the burden of proof is on those wishing to deny a possibility claim.[20]

3.3   A *reductio* against Structrualism

After offering *prima facie* counterexamples, I will now present an argument to the effect that Lewis is committed to reject Structuralism. My case that a given pair of worlds constituted a counterexample to structuralism has been conditional on an assumption: that some term of ordinary English belongs to O, or has a synonym in O. This assumption is hard to justify in a give case, since Lewis has left us in the dark about what O is. But the critic of the Replacement Argument has a way around that dialectical difficulty. For that argument has other premises that do tell us something about O that is in tension with 2). We can then argue against Structuralism by *reductio ad absurdum*.

That strategy immediately faces a problem: while the thesis of Humility, the conlusion of the Replacement Argument, is surprising, it is hardly absurd. Perhaps there are possible fundamental properties F and G that differ so subtly that we are in principle unable to know whether F or G is actually instantiated. Humility is not a suitable basis for a *reductio*. The key to the *reductio* strategy is realizing that the premises of Lewis's argument do not just entail Humility, but also another, truly implausible claim. The Replacement Argument proves too much.

The claim that I take to be absurd is the following:

8) We cannot entertain any proposition that does not supervene on fundamental structure.

In other words: we cannot entertain any proposition that is true in one of a pair of fundamentally isomorphic worlds and false in the other.

In one sense, 'to entertain A' means to deliberate about A without committing yet to either its truth or falsity. The sense in which the expression is used in 8) is weaker. If you have a propositional attitude towards A, you *eo ipso* entertain A.[21]

There is no reference to O in 8), and it thus can be assessed without knowing what O is. An implementation of the strategy now requires two things: first, showing that the Replacement Argument indeed commits Lewis to 8), and arguing that 8) is, if not absurd, at any rate false.

The second task has in effect been dispatched already. In 3.2, I tried to describe worlds that are fundamentally isomorphic yet qualitatively different, and thereby to express, and to make the reader entertain, propositions that are true in one but not the other. In so far as I succeeded, I have established that 8) is false. Which terms are in O, and which ones are not, is neither here nor there.

Thus there remains the first task: showing that the premises of Lewis's Replacement Argument entail 8). Of course, it does not follow from the premises 1), 2), 3) and 4) of my reconstruction in section 2. But that reconstruction was only as fine-grained as was needed then.

The premises that, jointly with Structuralism, entail 8) are hidden in a sub-argument for the expressibility premise 1):

1) If A is knowable, then it is entailed by a true proposition that is expressible in O.

We can reconstruct an argument for premise 1) as follows, where L_T is the language in which the final and complete theory T is formulated:

1a) If A is expressible in L_T, it is expressible in O.

1b) If A is entertainable, it is expressible in L_T.

1c) If A is knowable, it is entailed by a true entertainable proposition.

From 1a) and 1b), it follows that entertainable propositions are expressible in O. Together, 1a) - 1c) imply 1).

Crucially for present purposes, premises 1a), 1b) and 2) together imply that we can only entertain propositions that supervene on fundamental structure, and hence 8).

Premise 1c) ought to be uncontroversial. Knowledge that p implies belief that p, and the belief that p implies the entertainability of p, or at least of some something that entails p.[22] It remains for me to show that Lewis is committed to 1a) and 1b).

To see why premise 1a) holds, we need to take a closer look at the language L_T in which the hypothesized final and complete theory T of the world is formulated. The language O is just a fragment of L_T: in addition to O-terms, L_T also contains T-terms, or "theoretical" terms. In the framework Lewis deploys, T-terms are introduced, and implicitly defined, by the theory T.[23] In Lewis [1970, section IV], Lewis showed how implicit definitions can be turned into explicit ones. Explicit definitions are mere abbreviations: they do not increase the expressive power of the language.[24] To be sure, not each term of L_T needs to have a synonym in O. But while some L_T-expressible propositions might not find a succinct expression in O, they will find some expression, however cumbersome. Hence as a matter of logic, any proposition expressible in L_T is also expressible in O.

To see why premise 1b) holds, we need to recall what role the theory T plays in the Replacement Argument. T is meant to imply all truths that we can think or say. If it did not, then we should move to a more comprehensive theory, and try to run the Replacement Argument for it. Hence L_T must be able to express any proposition that we can express, in natural language or in scientific language. Indeed, the expressive power of L_T must not only match, but exceed the expressive power of our actual language. The argument is about limits to what we can know in principle, not about our present epistemic limits. Premise 1b) would thus not be warranted if L_T were an actual language used at a particular stage in history. All such languages have contingent limitations in their expressive power that humans might eventually overcome. That something is not expressible in an actual language at best entails that it fails to be known, not that it fails to be knowable. Clearly, Lewis allows that L_T is idealized, free from contingent limitations.

I thus conclude that a proponent of the Replacement Argument is committed to 8). (Though I will briefly reconsider the question in 4.1, when I respond to an objection.) As I noted, all the examples I gave in 3.2 are *prima facie* counterexamples to 8), quite independently of any assumptions about which terms do or do not appear in O. But I would like to strengthen my case against 8) further by an *ad hominem* move against Lewis: 8) implies that materialism, as defined by Lewis himself, is not entertainble. If it is not entertainable, then it cannot be believed, or disbelieved, or doubted, or argued about—a conclusion that some philosophers would have welcomed, but certainly not Lewis.

For Lewis, materialism is compatible with modal space including outlandish, spirit-ridden possible worlds (see Lewis [1983a], Lewis [1999, What Experience Teaches, pp. 274-5], and Lewis [1999, Reduction of Mind, pp. 291-3]), as long as they are not in the "inner sphere of possibility" around the actual world. According to his definition, a world is *materialistic* iff among the worlds in the inner sphere around it, no two differ without differing in the distribution of physical

properties.[25] The inner sphere of worlds around w consists of those worlds in which no fundamental properties alien to w exist.[26] If all fundamental properties in a world are physical, then it is materialistic. For the fundamental properties in any two worlds in the inner sphere are likewise physical, and hence these worlds do not differ without differing physically. Conversely, if a world is materialistic, then all its fundamental properties are physical.[27] So we can equivalently take materialism to be true in a world iff all fundamental properties in that world are physical.[28]

Surely, there can be physical and non-physical fundamental properties in the same category. For example, mass is a scalar-valued monadic physical property. In some world where dualism is true, pain perhaps is, or is realized by, a scalar-valued monadic non-physical property. It is now easy to see how the *reductio* is completed. A combinatorial principle will imply that there are two fundamentally isomorphic worlds such that materialism is true in one of them but not the other.[29] Thus materialism, as defined by Lewis, does not supervene on fundamental structure, and by 8) fails to be entertainable.[30]

The same sort of argument applies to many other contingent supervenience claims as well: they fail to be entertainable if 8) is true. Exceptions are provided only by claims to the effect that all actual fundamental properties and relations belong to a certain category. The Leibnizian doctrine that all relations supervene on monadic properties would not be susceptible to the argument, since monadic properties form their own category and thus cannot be replaced by non-monadic relations while preserving fundamental structure.[31]

3.4   Structuralism is unsupported

I have given reasons for rejecting Structuralism. Here I consider whether there are also reasons to accept it, at least for Lewis. Are there other principles, endorsed by him, that entail Structuralism? I argue that there are not.

In 3.1, I noted that Structuralism is false if some property constant in O denotes a fundamental property. In "Ramseyan Humility", Lewis is explicit that there are no such constants in O. He writes that "fundamental properties are not named in O except as occupants of roles" (p. 3). Presumably, a property is named *sans phrase* in O if it is denoted by an atomic term, and it is named as an occupant of a role if it is definable in O by the Ramsey-Carnap-Lewis method [Lewis, 1970]. I call this the "anonymity constraint":

Anonymity constraint No fundamental properties are denoted by atomic O-terms.[32]

Given a gloss of O as the non-theoretical part of the language, the anonymity constraint has some plausibility. Arguably, a term standing for a fundamental property needs to be a theoretical term.

   Lewis does not impose any further constraints on what sort of properties could be denoted by atomic terms of O. But does the anonymity constraint alone entail 2), the structuralist premise? No. Let F and G be fundamental properties, and suppose that 'P_1' denotes the conjunction of F and G, or $C(F, G)$ for short.[33] Further, suppose that 'P_2' denotes $C(F, NG)$ (where NG is the negation of G), and 'P_3' denotes $C(NF, G)$. Then an individual has F if and only if it satisfies '$P_1(x) \vee P_2(x)$', and an individual has G if and only if it satisfies '$P_1(x) \vee P_3(x)$'. Thus these two properties are named in O as the occupiers of the roles corresponding to these open sentences, but not named in any other way. Now consider the sentence '$\exists x\ P_1(x)$'. This sentence is true in a world in which some individual has both F and G, but false in a fundamentally isomorphic world in which G has been replaced by H. Thus Lewis's anonymity constraint does not guarantee the truth of 2).

   Are there plausible stronger constraints that do entail 2)? Does Structuralism follow from a principle in the same spirit as the anonymity constraint?

   A natural candidate is the following:

Strong anonymity constraint No Boolean combinations of fundamental properties and relations are denoted by atomic O-terms.

It is easy to see that this constraint does not entail 2) either. Suppose a pair of individuals satisfies '$P_4(x, y)$' if and only if the first has F and the second has G. The term 'P_4' denotes a two-place relation which is not a Boolean combination of monadic fundamental properties like F and G. Then again, O can distinguish between fundamentally isomorphic worlds: '$\exists xy P_4(x, y)$' is true in world with two individuals, one of which has F and one of which has G, but false in another world where G is replaced by H.

   The strong anonymity constraint only bans properties definable from fundamental ones by the application of Boolean operators. We can formulate an even stronger condition:

Ultrastrong anonymity constraint If a property or relation is definable from the class of fundamental properties and relations using identity and (finitary and infinitary) quantifiers and Boolean operators, it is not denoted by an atomic O-term.[34]

However, this constraint does not imply 2) either. For a counterexample, consider two worlds w and w' of two-way infinite recurrence which are fundamentally isomorphic, but where F is instantiated in w wherever F' is in w'. Suppose G is a non-fundamental property such that for one epoch in w, G is had only by individuals in that epoch in w, and by no individuals in w'. Since not even infinitary logical resources would enable us to define G from fundamental properties and relations, the ultrastrong anonymity constraint allows G to be denoted by

an atomic O-term 'P'. Then the sentence '$\exists x P(x)$' expresses a proposition that fails to supervene on fundamental structure. (We do not have to assume anything about how G is distributed in worlds other than w and w'. It may apply to everything in the actual world, and to nothing in worlds fundamentally isomorphic to it.)

A supporter of Structuralism may try to rule out such counterexamples. She could deny that there are highly "unnatural" properties like G, insisting that properties are relatively sparse. But this strategy is unpromising, for the ultra-strong anonymity constraint is independently implausible. The only properties that it allows to be denoted by atomic O-terms are of a very special sort: they are *symmetry-breakers*. Some worlds have symmetries in the distribution of their fundamental properties: they may be spatially symmetric, or there may be two-way eternal recurrence. If w is such a symmetric world, there are fundamental isomorphisms from w to itself that do not map every individual to itself. A property F is a symmetry-breaker iff there is a world w, an individual x in w, and a fundamental isomorphism μ from w to itself such that x has F and $\mu(x)$ does not. The only properties that cannot be defined from the fundamental properties and relations using infinitary resources are symmetry-breakers.[35]

But the claim that only symmetry-breakers are denoted by an atomic O-term presents a dilemma. Either no properties at all are denoted by atomic O-terms, or some are. The first horn is clearly unacceptable, for it entails that O consists of logical vocabulary only.[36] We need to remember that O is the language in which our complete theory is formulated. Everything we can say is expressible in O, and we are clearly not restricted to logical vocabulary. The second horn leaves us with a mystery in foundational semantics. Plausibly, symmetry-breakers are among those properties that we cannot pick out, either by atomic or by defined terms. How can we denote any property that applies to horses in epoch e but not to horses in epoch e' in some world of two-way eternal recurrence? If atomic terms can achieve the feat of picking them out, why do they not manage the apparently lesser feat of picking out properties that do not break symmetries?

The ultrastrong anonymity constraint is most plausible under the assumption that the actual world displays symmetries in the distribution of fundamental properties. Perhaps our atomic O-terms apply to things in our vicinity, but not to things in symmetrically related, but spatially or temporally distant parts of our world. But then the inference to Structuralism can only be secured by further *ad hoc* assumptions. Suppose that in the actual world, individual x is in the extension of atomic O-terms 'P' and 'Q', while none of the individuals to which it is symmetrically related is. Let w be fundamentally isomorphic to @, and let X be the class of those individuals in the domain of w which are mapped to x by some fundamental isomorphism from w to @. Then Structuralism implies that there is exactly one member y of X which is in the extension of 'P' in w; and

that that very y is also the only member of X that is in the extension of 'Q' in w. But which member of X is this "favourite counterpart" of x? Any answer seems entirely arbitrary.

As this discussion indicates, there is little prospect of deriving Structuralism from some independently plausible strengthened version of the anonymity constraint.

Should we consider Structuralism to be a consequence of principles which Lewis articulates in other papers, and which might work as a background assumption in "Ramseyan Humility"? Of course, there is no space here to conclusively establish the negative existential claim that there is no such principle. Nonetheless, I want to mention an example of a claim from which Structuralism might be thought to follow, but does not, as a moment's reflection shows.

If we are speaking loosely, we might express Lewis's well-known doctrine of Humean supervenience by saying that "everything supervenes on fundamental structure." But Structuralism as understood here emphatically does not follow from Humean supervenience. Roughly, the latter is the thesis that if two worlds have the same fundamental properties instantiated in corresponding spacetime-points, they do not display any qualitative differences [Lewis, 1986b].[37] Evidently, it does not follow that if two worlds have different fundamental properties, but arranged in the same way, they do not display any qualitative differences. An analogy: the claim that no two mosaics differ in their aesthetic properties if they consist of the same arrangement of stones is very different, and much weaker, than the claim that no two mosaics differ in their aesthetic properties if they have isomorphic arrangements of stones, even if blue stones have been swapped throughout for red ones and vice versa.[38]

3.5   Does Contingent Structuralism help?

In 3.1, I noted a *caveat* regarding my reconstruction of the Replacement Argument. Strictly speaking, Lewis only needs that the following condition is satisfied by the actual world, not that it is satisfied by every possible world (T_w is the true and complete theory of world w):

2') If σ realizes T_w in w, then every replacement σ' of σ realizes T_w in $\sigma'w$.

I showed that 2), i.e. Structuralism, follows from the assumption that 2') is satisfied by every world, and went on to argue against that premise. Using the same reasoning as in 3.1, we can show that the following weaker claim follows from the assumption that 2') is satisfied by the actual world:

Contingent Structuralism If A is expressible in O, then it has the same truth-value in all worlds fundamentally isomorphic to the actual world.

The argument to the conclusion 7) then goes through if we substitute all occurences of "supervenes on fundamental structure" with "has the same truthvalue in all worlds fundamentally isomorphic to the actual world".

I will use the convenient name "Contingent Structuralism" even though it is not quite accurate. The above claim does not imply that anything is contingent; it merely fails to imply that it is necessary that the fundamental structure determines the complete O-description.

In 3.1, I argued that we should not assume that the actual world satisfies 2′) unless we are assuming that all worlds do. In other words, we should accept Contingent Structuralism only qua consequence of Structuralism, if at all. Otherwise, we are claiming that the actual world is in some basic respect special. Lewis himself has warned us from such actualocentrism. A version of the Replacement Argument that appeals to Contingent Structuralism as an unsupported premise is thus dialectically weak.

Nonetheless, it is worth asking whether we have good reason to think that Contingent Structuralism is false. I want to argue that the case against Structuralism undermines Contingent Structuralism as well.

The discussion in 3.4 carries over without change. The anonymity constraint and its stronger cousins all appear to be either extremely implausible, or else do not entail Contingent Structuralism.

The discussion in 3.3 carries over with only minor change. Given a sufficient supply of both physical and non-physical fundamental properties and relations in the actually instantiated categories, we can argue that there are both materialistic and non-materialistic worlds among those that are fundamentally isomorphic to the actual world. Thus Contingent Structuralism, together with 1a) and 1b), also implies that materialism is not entertainable. Again, the retreat to Contingent Structuralism does not help.

However, some comment is required about how the discussion in 3.2 applies to Contingent Structuralism. The counterexamples to Structuralism involved "toy" worlds with a very simple structure. It is of course much harder to describe pairs of worlds which share the enormous complexity of the actual world.

Nonetheless, as I noted in 3.2, I do not find it in the least plausible that O-differences among fundamentally isomorphic should be restricted to worlds of low complexity. Indeed, it seems to me that ideas from the previous examples can be used to describe a world w that is fundamentally isomorphic but O-discernible from the actual world. To adapt example 1: For any actual individuals having a determinate of mass, there are individuals in w having a corresponding determinate fundamental non-physical property; and likewise for other actual fundamental properties. Then sentences such as "there is a material thing" and "some physical property is instantiated" express propositions that are true in @ and false in w. If either of the terms 'material' or 'physical' belongs to the

language O, or has a synonym in O, then these propositions can be expressed in O, and Contingent Structuralism is false.

One might worry whether individuals that do not have physical properties can stand in spatiotemporal relations. Arguably, angels would fit that description. If their possibility is not accepted, one would have to introduce non-spatiotemporal fundamental relations as well, as in example 4. This confirms what I said before: more complex worlds require a more complex discussion. Nonetheless, there is little hope for the proponent of the Replacement Argument that Contingent Structuralism can be salvaged from the wreckage of Structuralism.

4 Objections considered

So far, I have been on the attack, criticizing the Replacement Argument by undermining its second premise. In this section, I switch to defense. I consider three objections to the effect that the Replacement Argument can be reconstructed in such a way that it is not vulnerable to my criticism. In response, I argue that it does not fare better on these alternative reconstructions.

4.1 Unknowability without Inexpressibility?

The *reductio* argument against Structuralism depends on 1a)-1c). But surely, someone not committed to those premises might still advance a variant of the Replacement Argument? She might support premise 1) in some other way. Even expressible propositions may be unknowable, after all. Another necessary condition for the knowability of p is that we could have justification to believe it, and we can arguably only have such justification if we have evidence for p. Perhaps there could not be any evidence for those propositions not expressible in O. This line of thought suggests the following alternative sub-argument for premise 1):

1a′) If A is knowable, it is entailed by a true proposition that we could have evidence for.

1b′) If A is not expressible in O, we could not have evidence for it.

Given a language O, we can distinguish two characters who may run the Replacement Argument, whom I call the "expressive skeptic" and the "evidential skeptic." The expressive skeptic denies that we can express, or entertain, any propositions that are not expressed by a sentence of O. Anything not translatable into that idiom is unintelligible to her. The evidential skeptic grants that we can think and speculate about many questions that cannot be asked in O. But she insists that such speculation is entirely idle, since there could not be any evidence that favoured one answer over another one.

My *reductio* from 3.3 works only against an expressive skeptic, who appeals to premises 1a)-1c) in addition to 2)-4). It has no force against an evidential skeptic. This raises the question whether the Replacement Argument is effective in the hands of the evidential skeptic.

The answer is negative, for the reasons laid out in Schaffer [2005]. If we can express different hypotheses about what the actual fundamental properties are, but O cannot express them, then so much the worse for the claim that only O-expressible propositions could bear on whether we should believe these hypotheses. Schaffer points out that the evidential skeptic's Replacement Argument is just another skeptical argument of the kind familiar from the tradition. He then argues that any epistemological theory that has a response to skepticism can likewise disarm that particular argument.[39]

What about the exegetical question: is the author of "Ramseyan Humility" an expressive skeptic or an evidential skeptic? I already argued in 3.3 that he is an expressive skeptic. Lewis explicitly states that all terms in L_T are either also in O or defined in terms of O by the theory T. Hence he is committed to accept 1a), i.e. that the expressive power of L_T does not exceed the expressive power of O. Moreover, he explicitly takes T to be a complete theory. The intended sense of completeness is that it answers all the questions we can think of. (It is not meant to be complete in the sense of implying every true proposition; otherwise, its completeness would be incompatible with 6).) Thus Lewis is committed to 1b) as well.

When taken in isolation, some passages in Lewis [forthcoming] may suggest that Lewis is an evidential skeptic:

> O-language, we assumed, is rich enough to express all possible observations. Therefore any predictive success for T is equally a predictive success for the Ramsey sentence of T. Since the evidence for T consists of its record of predictive success, there is no way to gain evidence for T that is not equally evidence for the Ramsey sentence.
> ...
> There is indeed a true contingent proposition about which of the possible realizations is actual, but we can never gain evidence for this proposition, and so can never know it. If there are multiple possible realizations, Humility follows. (p. 5/6)

But Lewis also makes clear that the allegedly unknowable propositions are not entertainable:

> We cannot answer the question: which property occupies that role? But worse: not only can we not answer that question, we can't even ask it. (p. 16)

An answer to that question would be the sort of proposition that the thesis of Humility claims to be unknowable: one that is true only in worlds in where some fundamental property F is instantiated (and plays a given role), but not in worlds where F is replaced by a distinct G.

The passages above do not show that Lewis is an evidential skeptic. We should note that anyone who is committed to 1a)-1c) will also accept 1a$'$) and 1b$'$), and is thus entitled to use the sub-argument for 1) presented in this section. For premise 1a$'$) is uncontroversial if we take 'evidence' broad enough. Moreover, for a fixed O, 1b$'$) seems weaker than the conjunction of 1a) and 1b)—we could not have evidence for a proposition that is not entertainable, presumably. Given that an expressive skeptic has the stronger commitments, she can still assent to these premises. Thus the hypothesis that the author of "Ramseyan Humility" is an expressive skeptic is not at odds with what he says in these passages.

4.2   Replacing the structuralist premise?

In my reconstruction, Lewis relies on Structuralism, the thesis that any two fundamentally isomorphic worlds are O-indiscernible. However, a weaker premise in place of Structuralism would still yield a valid argument for Humility: that among those non-actual worlds that are fundamentally isomorphic to the actual world, *some* are O-indiscernible from it.

Here is the idea. What O-expressible propositions are true at a world depends, in general, on two things: what fundamental structure the world exhibits, and what fundamental properties and relations are at the nodes of that structure. Thus replacing a fundamental property in a structure sometimes results in an O-difference. But it need not always make a difference, for all that has been said. The new argument relies on the claim that if a property can be replaced at all, there are some possible replacements that leave the O-truths unaffected; instead of a universal premise, it uses an existential premise.

For comparison with the old one, I present this new argument in deductive form as well. The new, weaker premise 2*) is here formulated in terms of entailment and supervenience on fundamental structure. All the other lines are the same as in the reconstruction in section 2:

1) If a proposition is knowable, then it is entailed by a true proposition that is expressible in O.

2*) If the existence of F is not entailed by any true proposition that supervenes on fundamental structure, it is not entailed by a true O-expressible proposition either.

3) If not every member of category C is actual and a true proposition entails that $F \in C$ exists, it does not supervene on fundamental structure.

4) Not every member of category C is actual.

5) If a true proposition entails that $F \in C$ exists, it does not supervene on fundamental structure. (from 3) and 4))

6) If a true proposition entails that $F \in C$ exists, it is not expressible in O. (from 2*) and 5))

7) If a proposition entails that $F \in C$ exists, it is not knowable. (from 1) and 6))

Premise 2*) can be paraphrased as follows: if there are fundamentally isomorphic worlds that differ with respect to the existence of F, then at least two of them are O-indiscernible.

How could one go about arguing for this existential claim? In the analogy, three strategies come to mind. First, the *way of example*: finding two leaves that prove to be exact duplicates under microscopic scrutiny. Second, the *way of grouping*: arguing that all oak tree leaves must be qualitatively identical, and that there are several oak tree leaves in the garden. Third, the *way of counting*: arguing that given the ratio of the number of leaves in the garden to the number of possible qualitative variations among leaves, it is certain, or at least probable, that some leaves are qualitatively identical.

Although valid, the new Replacement Argument is dialectically ineffective as it stands. Premise 2*) says, roughly, that if combinatorialism holds, quidditism does too. But in the context of the argument, the transition from combinatorialism to quidditism is at issue. Suppose you are a combinatorialist, but cannot see why your view is committed to any claim about expressibility, or about the language O. In particular, you wonder why it is committed to the version of quidditism expressed by 6). Why should you be swayed to endorse quidditism by an argument that appeals to 2*)?

Premise 2*) is not a natural fall-back position once 2) is given up. An analogy: Suppose I think that all leaves in a garden are exactly qualitatively identical. Perhaps I hold this as a default assumption; or I accept it as a consequence of the claim that all features of a leaf are determined by its genetic make-up, and the claim that all leaves are genetically the same. If I am then presented with examples of qualitatively distinct leaves, the existential claim that there is at least one pair of qualitatively identical leaves is not automatically a natural fall-back position for me. A special argument would be needed.

How could one go about arguing for this existential claim? In the analogy, three strategies come to mind. First, the *way of example*: finding two leaves that prove to be exact duplicates under microscopic scrutiny. Second, the *way of grouping*: arguing that all oak tree leaves must be qualitatively identical, and that there are several oak tree leaves in the garden. Third, the *way of counting*: arguing that given the ratio of the number of leaves in the garden to the number of possible qualitative variations among leaves, it is certain, or at least probable, that some leaves are qualitatively identical.

In the case of the new Replacement Argument, there is likewise a special argument needed for premise 2*), if it is not accepted qua consequence of 2). Can any of the three strategies be deployed successfully to argue for 2*)? The way of examples might seem promising: does permutation of negative charge and positive charge not result in an O-indiscernible world; and is there not an (arguably

alien) property, negative mass, that could take the place of mass without giving rise to O-differences?

This suggestion seems to presuppose that we can refer to mass and charge, and that we know of them that they are fundamental. It may thus threaten to deliver a Phyrric victory in the battle for premise 2*), which results in the loss of the war to save the argument. But even if if it is granted that such presuppositions can be avoided, the suggestion still faces a dilemma.

Is there one fundamental property of electric charge which takes both positive and negative real values? Or are positive and negative electric charge distinct fundamental properties that both take positive real values, and that nomologically exclude each other in the actual world?[40]

On the first horn, we do not have a pertinent example: rather than a permutation or replacement of a fundamental property that leaves O-truths unaffected, we would only have a permutation (in the case of charge) or replacement (in the case of mass) of *values* of one and the same fundamental properties that leaves O-truths unaffected. The possibility of permutations and replacements of the latter sort is much less striking, and does not lead to Humility as characterized by Lewis.

On the second horn, positive charge and negative charge are distinct fundamental properties in the same category. A combinatorial principle such as the one invoked by Lewis will thus apply to them. Hence there will be pairs of worlds w and w' of the following sort: in w, everything has two units of positive charge and one unit of negative charge; and w' is exactly like w except that everything has two units of negative charge and one unit of positive charge. The worlds w and w' are fundamentally isomorphic. Are the same O-sentences true in them? I, for one, am finding it extremely hard to develop intuitions about this. Perhaps w and w' are O-indiscernible, and perhaps they are not. But at any rate, the claim that they are O-indiscernible cannot be supported by pointing out that positive and negative charge play symmetrical roles in the fundamental laws of our world. For on the hypothesis being considered, these laws are entirely contingent, and will fail in w or w'.

Does the way of grouping fare better? Presumably, the class of physical properties is the most promising candidate for a sub-class of the fundamental properties for which a restricted version of Structuralism holds. But are the physical fundamental properties really such that their exchange in a structure does not result in O-differences? Consider again a world where positive charge is instantiated wherever mass is in the actual world, and *vice versa*, and all other fundamental properties are distributed as they actually are. Such a world is presumably incompatible with actual laws, but combinatorialism implies its metaphysical possibility. Is there a desk in the region corresponding to the region that my desk occupies? Is it heavy, as is my desk? I have no idea what features

the thing in that region has. Why should I think that it is heavy? If it is not, then the way of grouping fails for the class of physical properties.

Philosophers have made various proposals about what the distinguishing marks of physical as opposed to non-physical fundamental properties are. But as far as I know, it has not been suggested that physical properties can be permuted with and can replace other physical properties without making a difference to the O-description of a world; and that replacing them with non-physical properties, in contrast, may make for an O-difference. The strategy asks the distinction between physical and non-physical properties to bear more weight than it can carry. If premise 2*) is to be supported by the way of grouping, we would need to be told more about physical properties than we have been so far.

Finally, the way of counting. The idea here is that there are so many different fundamental properties that there are not enough O-roles to distinguish them all. To be sure, actual languages have only limited resources, and would not be able to distinguish all fundamentally isomorphic worlds, if there are sufficiently many. But it is clear that Lewis allowed us to idealise away from that sort of limitation of O. Hence the way of counting does not seem promising.

To emphasize again, I am not claiming that 2*) is false, nor that no good argument in its favour could be given. I consider it to be an interesting and difficult question whether 2*) is true, a question best approached with a substantive metaphysical conception of what fundamental properties are. What I insist on here is merely that no argument for 2*) is implicit in "Ramseyan Humility", or indeed in other publications of Lewis.

4.3   Two-Dimensionalism to the Rescue?

My reconstruction of the Replacement Argument presupposes that there is a functional relation of expression between sentences and propositions. I often talk about "the proposition expressed" by a sentence, or about propositions "expressible in O." It may be objected that this is problematic. Lewis himself appeals to what is called "two-dimensionalism" in a later stage of the paper, and according to two-dimensionalism, there are two salient expression relations between sentences and propositions, which we may call "primarily expressing" and "secondarily expressing." Does this distinction undermine my criticism of the Replacement Argument?

Lewis invokes two-dimensionalism in response to the following objection. Let $T(P_1, ..., P_n)$ be the true and complete theory of @, in a given (finite or infinite) axiomatization, and grant that T's Ramsey-sentence $T^R =$

$$\exists X_1...\exists X_n T(X_1, ..., X_n)$$

is likewise true in worlds that are fundamentally isomorphic with @. Then

$$\exists X_1...\exists X_n[T(X_1, ..., X_n)\wedge \text{Actually } T(X_1, ..., X_n)]$$

is false in worlds where fundamental properties have been been permuted or replaced relative to the actual world. Indeed, that sentence expresses a maximally strong proposition which is true only in @ (and other worlds indiscernible from it, if there are any). If it belongs to O, the structuralist premise is false.

Lewis responds by distinguishing ways in which a proposition is expressed by a sentence. Instead of claiming that we do not have sentences that express an answer to the question which fundamental properties are actual, he claims that we have no sentences "that express those alternative answer-propositions, and do so in such a way that we can know which sentence expresses which proposition" [Lewis, forthcoming, p. 15]). A footnote makes clear that he is appealing to two-dimensional semantics. For him, it is the primary intension that is expressed "in such a way that we can know which sentence expresses which proposition."[41] 'Expressible', as I have used it, ought to be understood as "expressible qua primary proposition of a sentence or thought," and 2) should be understood as 2'):

2') If a proposition is the primary intension of a sentence of O, it supervenes on fundamental structure.

This move foils the above objection, since the proposition that is true in @ but not in the fundamentally isomorphic worlds is only expressed as the *secondary* intension of the above sentence. Adding a rigidifying operator such as 'Actually' makes no difference at all to the primary proposition associated with a sentence.[42]

Consider "Things are as they actually are," or "Every proposition is true if and only if it is actually true." The primary intensions of these sentences are necessary, while the secondary intensions are true and maximally strong. There is a good sense in which these sentences are uninformative and do not imply answers to questions about our world. Lewis seems at least *prima facie* justified in insisting that the primary intension is the relevant one for the Replacement Argument.

Does the substitution of 2') for 2) blunt my objections? The sentences I suggested as potential counterexamples are "there is a material thing", "some physical property is instantiated", "something is in pain", "something is red", and "two things are apart from each other, but close". None of them involves a rigidifying operator. Thus it appears that replacing 2) by 2') is of no help. But a proponent of the Replacement Argument might argue that the primary intension of these sentences does after all diverge from the secondary intension, and that only the latter differs between isomorphic worlds. On one strategy, she would posit hidden structure: the logical form of the above sentences involves rigidfiying operators. Perhaps 'pain' is to be understood as a rigidified definite description, roughly as 'state actually playing role R', where R would need to

be specified. On another strategy, the proponent of the Replacement Argument would generalize two-dimensionalism: the primary and secondary intension may diverge even in the absence of any overt or hidden rigidifying operators. I will argue that neither strategy promises to rescue the argument.

Pursuing the hidden-structure strategy in effect amounts to redrawing the boundary between the old and the new terms in the language. Lewis seems to tell us that O may contain all but the most obviously theoretical terms, i.e. those that pick out fundamental properties, and that no assumptions are made about its semantics. But the hidden-structure move is effectively treating the terms used in my examples as theoretical terms, in the sense of terms that are introduced by description via the Ramsey-Carnap-Lewis method. It thus relies on strong claims about the structure of O. Moreover, some of the claims needed to defuse my examples are highly implausible. Notoriously, pain does not seem to be definable as the actual realizer of a role.

The generalization strategy could be developed in many different ways. We would need to be told what the primary intension of a sentence S is, if it is not just what is expressed by S', where S' results from deleting rigidifying operators in S. Even if this could be done, it might fail to rescue the Replacement Argument for at least three reasons. First, the primary and secondary intensions might not diverge for all terms, and counterexamples to the structuralist premise could be constructed from those terms for which they coincide.[43] Secondly, primary intensions may themselves distinguish between isomorphic worlds, and hence not be of any use in the defence of Structuralism. Thirdly, the Replacement Argument might fail even if primary intensions are characterized in such a way that 2′) is satisfied, and thus deserve to be called "structural intensions." For once premise 2) has been replaced by 2′), the argument only goes through with a compensating change in the first premise, from 1) to 1′):

1′) If A is knowable, then it is entailed by a true proposition that is primarily expressible in O.

It might be plausible that knowledge must be expressible, but why should it be expressible qua primary intension of a sentence? Here is an ineffective argument for that claim: Since structural intensions can never differ between epistemically indiscernible worlds, and since fundamentally isomorphic worlds are epistemically indiscernible, structural intensions can never differ between fundamentally isomorphic worlds. It is clear why such a defense would not be of any use to the proponent of the Replacement Argument. The claim that all fundamentally isomorphic worlds are epistemically indiscernible is not pre-theoretically obvious; I, for one, find it *prima facie* little more attractive than the claim that isomorphic mosaics look the same. Moreover, the claim is rather close to the thesis of Humility. If there were independent support for it, it would threaten to make the Replacement Argument redundant. If there is no independent support, then the

Replacement Argument might be described as circular, or begging the question, or failing to transmit warrant, depending on what your favorite diagnostic tool is. In any case, it would be dialectically ineffective.

I conclude that a two-dimensionalist defence of the Replacement Argument is undermotivated. Of course, I cannot show that no modification of that argument succeeds. But the prospects are dim. In contrast, I do not take Humility, the conclusion, to be too implausible. What I hope to have established here is that if we accept Humility, it ought not to be on the basis of Lewis's Replacement Argument.

Notes

[1]Alternatively, the content of the theory is given by a variant of the Ramsey sentence that replaces existential quantifiers with uniqueness quantifiers ($\exists!$). I will disregard subtleties that arise if a theory has more than one realization in the same world.

[2]Lewis borrows 'humility' as a philosophical term of art from Langton [1998], who uses it to label an ignorance claim she attributes to Kant. Psillos [2006] was independently inspired by Langton to coin the expression "Ramseyan Humility" for a similar thesis he attributes to Ramsey.

[3]For example, David Armstrong's combinatorial theory of possibility [Armstrong, 1989] would license an argument from permutation, but not from replacement.

[4]For critical examination of that thesis, see Schaffer [2007].

[5]It is faithful, but not maximally fine-grained. There are sub-arguments for some of the premises that I skip over here. This will become relevant in 3.3 below.

[6]An item of terminological book-keeping: in my usage, a false proposition A is not knowable, even if there is a possible world where it is not just true, but known by subjects in that world.

[7]Unfortunately, the term "quidditism" has different uses; and in "Ramseyan Humility", Lewis himself uses it in a non-standard way. The claim he calls "quidditism" is already a trivial consequence of the claim that I call "combinatorialism".

[8]Here as elsewhere in the paper, I speak of properties rather than properties and relations, for the sake of simplicity. The definition of a fundamental isomorphism given here can straightforwardly be extended to relations.

If we are using the framework of *centered* possible worlds, then a fundamental isomorphism is also required to map the center of w to the center of w'.

[9]I here ignore the epistemic possibility, mentioned by Lewis in footnote 4, that there is no minimal supervenience base.

[10]For Lewis, a fundamental property is an idler at a world if it does "not play an active part in the workings of nature" (p. 3) of that world.

[11]I am here ignoring irrelevant complications due to idlers.

[12]"What we need to know about the Ramsey sentence is that it logically implies exactly the O-language sentences that are ... logically implied by the postulate of T" [Lewis, forthcoming, p.4]. The postulate of T is what I am just calling 'T'. Since it is complete, any sentence is either logically entailed or logically incompatible with it.

[13]The same comments apply, *mutatis mutandis*, to Lewis's statement of the permutation argument: "Suppose we have the actual realization of T. Maybe some members of the n-tuple that realizes T are not fundamental properties, or maybe some belong to single-membered categories. Hold those ones fixed. Permute the rest within their categories to obtain a new n-tuple. It too would realize T." (p. 5)

[14]'$P(x)$' here abbreviates 'x instantiates P'. I take O to be a first-order language with singular terms for both individuals and properties, and with instantiation predicates (an $n+1$-place predicate for n-adic relations) as part of its logical vocabulary.

[15]The assumption that 'in pain' or a synonym is in O is entirely compatible with the anonymity constraint, discussed in 3.4. The term does not "name" F; it applies to properties other than F as well. Hence the assumption carries no commitment to the so-called "identification thesis" that Lewis rejects in the last section of Lewis [forthcoming].

[16]Chalmers [2006, pp. 79-80] tentatively concludes that worlds with such edenic color properties are possible.

[17]It could be objected that on Lewis's theory of possible worlds, it is arguably necessary that there is spacetime structure, such that there could not be a world such as w. However, Lewis is explicit that his replacement argument is not supposed to be hostage to his views

about possible worlds. Indeed, the argument would be of much more limited interest if it were presupposing these views.

[18]In fact, Lewis's anonymity constraint, discussed in 3.4, does not look all that plausible for fundamental spatial, temporal, or spatiotemporal relations. I will not develop that objection here, though.

[19]This would be entirely compatible with the supervenience of causation on fundamental properties and relations, a point that will be further explained in 3.4. In particular, it is also compatible with Lewis's accounts of causation in terms of counterfactuals and indirectly in terms of laws of nature. Fundamentally isomorphic worlds may not have isomorphic laws, for the comparison of candidate systems in terms of strength may be sensitive to what the fundamental properties are. Likewise, similarity orderings among worlds, which determine which counterfactuals are true, may be sensitive to what the fundamental properties are.

[20]To echo Zimmerman [1999, p. 213]: "I am fortunate to have Lewis for challenger ...: Lewis ... is willing to take the fact that something seems plainly possible as weighty evidence for its actually *being* possible—not just 'epistemically possible', i.e. true for all we know right now."

[21]Of course, it is a vexed question under what conditions one counts as having an attitude to a proposition. Clearly, not every relation to a proposition qualifies, but only fairly natural relations of the right kind. I will not attempt an analysis of what it is to have an attitude towards a proposition. While hard to analyse, the notion is understood reasonably well.

One can replace 'entertainable' by 'expressible' if preferred. It is an interesting question whether some propositions are entertainable but not expressible, but we can abstract from it for the purpose of the argument.

[22]Since every proposition entails itself, 1c) is weaker than the claim that every knowable proposition is entertainable.

[23]"The language of T contains *T-terms*: theoretical terms implicitly defined by T. And there is all the rest of our language, call it *O*-language." (p. 3)

[24]Sometimes, this is expressed by saying that definitions need to satisfy an eliminability constraint. For discussion of logical features of definitions, see for example Suppes [1957] and Belnap [1993].

[25]This definition takes 'being physical' as a term for a second-order property that is primitive, i.e. unanalysed. Of course, Lewis does have things to say about it.

[26]Maybe not even all of those worlds are in the inner sphere; see Lewis [1994, section 1]. That issue does not bear on the present discussion.

[27]Suppose a world w has a non-physical fundamental property F. By a combinatorial principle, there is a world that differs from w in the distribution of F but not in the distribution of physical fundamental properties. Hence w is not materialistic.

[28]Lewis himself uses a formulation of that sort in Lewis [1999, Reduction of Mind, p.292]: "[A]ll fundamental properties and relations that actually occur are physical. This is the thesis of materialism."

[29]Since combinatorialism needs to be appealed to here, could the argument be construed as a *reductio* of combinatorialism rather than Structuralism? In principle, this is a theoretical option. But it is a very unattractive one, because the *reductio* requires something far weaker than a general combinatorial principle for fundamental properties. What is needed is only that physical and non-physical fundamental properties can be instantiated in worlds with the same fundamental structure, which is a weak combinatorial principle indeed.

[30]An objection: This argument only goes through if it is assumed that the second-order-property of being physical is essential to every property that has it, i.e. that no property can be physical in one world and non-physical in another. But this is a fair assumption to make, since Lewis's definition of materialism is inadequate without it.

For suppose α is the class of actual fundamental properties, and that all members of α are physical in @. Then intuitively, materialism should be true. But suppose there is a world w where some member G of α is non-physical, and where no fundamental property not in α is instantiated. By combinatorialism, there are many worlds that differ from w only in the distribution of G, but not otherwise (in particular, not in whether the other fundamental properties are physical or not). Surely, there is at least one world w' where G is likewise non-physical. Then w and w' are worlds without alien properties that differ without differing physically, and hence Lewis's definition misclassifies @ as non-materialistic.

[31] Whether Lewis's doctrine of Humean supervenience is susceptible is unclear. Perhaps a case can be made that properties instantiated by point-sized things, or by mereological atoms, form their own category.

[32] The last section of "Ramseyan Humility," entitled "Humility about Qualia," is a partial defence of the anonymity constraint.

[33] In my metalanguage, I use 'C' and 'N' as functional expressions applying to properties names; '$C(F, G)$' denotes the conjunction of F and G, and 'NF' the negation of F. I use these symbols instead of more familiar negation and conjunction symbols to distinguish these functors from the sentential connectives which are present in O.

[34] Or better, it is not denoted by an atomic *non-logical* O-term. The identity relation is trivially definable from the class of fundamental properties and relations using the identity predicate.

[35] Glanzberg [2001] discusses the great power of infinitary logics in defining supervenient properties. See also Stalnaker [1996].

[36] I am assuming here that the the non-logical terms of O, if there are any, do denote. Likewise, I am assuming in this paper that O does not contain names of individuals. Without that assumption, Lewis argument is even more vulnerable.

[37] Lewis restricts the quantification to worlds in the "inner sphere of possibility," where there are no alien fundamental properties. But even without this restriction, Humean supervenience does not imply structuralism.

[38] It goes without saying that Structuralism is not implied by any combinatorial principles about fundamental properties. That O has predicates such as P_1 to P_4 is entirely compatible with F, G, and H satisfying combinatorial principles.

[39] Plainly, though, the answers he surveys are not answers to the expressive skeptic's Replacement Argument, since they rely on the entertainability of the propositions whose knowability is in question.

[40] In Lewis [1986a, pp. 154-5], Lewis suggests that we should be be agnostic with respect to these two options, and argued that it is a problem for the linguistic ersatzer that she cannot afford such agnosticim.

[41] "Primary intension" is David Chalmers' term, whose Chalmers [1996, pp. 52-71] is a good introduction to two-dimensionalism.

[42] This follows immediately from the semantic clause: 'Actually S' is true at $\langle v, w \rangle$ iff S is true at $\langle v, v \rangle$.

[43] On Chalmers' account [Chalmers, 2004, p.191], there are semantically neutral terms, whose primary and secondary intensions coincide (ignoring the fact that primary intensions contain centered rather than uncentered worlds). They are used in the canonical description of scenarios (roughly, ersatz possible worlds).

References

David M. Armstrong. *A Combinatorial Theory of Possibility*. Cambridge University Press, 1989.

Nuel Belnap. On rigorous definitions. *Philosophical Studies*, 72:115–146, 1993.

David J. Chalmers. *The Conscious Mind. In Search of a Fundamental Theory*. OUP, 1996.

David J. Chalmers. Perception and the Fall from Eden. In Tamar Szabó Gendler and John Hawthorne, editors, *Perceptual experience*, pages 49–125. Oxford University Press, 2006.

David J. Chalmers. Epistemic two-dimensional semantics. *Philosophical Studies*, 118:153–226, 2004.

Michael Glanzberg. Supervenience and infinitary logic. *Nous*, 35:419–439, 2001.

John Hawthorne. Intrinsic properties and natural relations. *Philosophy and Phenomenological Research*, LXIII (2):399–404, 2001.

Frank Jackson and Graham Priest, editors. *Lewisian Themes. The Philosophy of David K. Lewis*. Oxford University Press, 2004.

Rae Langton. *Kantian Humility*. OUP, 1998.

Rae Langton. Elusive knowledge of things in themselves. *Australasian Journal of Philosophy 82*, pages 129–136, 2004.

David Lewis. How to Define Theoretical Terms. *Journal of Philosophy*, 67: 427–426, 1970. Reprinted in Lewis 1983b (pp. 78-95).

David Lewis. New work for a theory of universals. *Australasian Journal of Philosophy*, 61:343–377, 1983a. Reprinted in Lewis 1999 (pp. 8-55).

David Lewis. Putnam's paradox. pages 221–236, 1984. Reprinted in Lewis 1999 (pp. 56-77).

David Lewis. Humean supervenience debugged. *Mind*, 103:473–490, 1994. Reprinted in Lewis 1999.

David Lewis. *On the Plurality of Worlds*. Blackwell, 1986a.

David Lewis. Ramseyan Humility. In David Braddon-Mitchell and Robert Nola, editors, *Naturalism and Analysis*. MIT Press, forthcoming.

David Lewis. Introduction. In Lewis 1986c. 1986b.

David Lewis. *Philosophical Papers*, volume I. Oxford University Press, Oxford, 1983b.

David Lewis. *Philosophical Papers*, volume II. Oxford University Press, Oxford, 1986c.

David Lewis. *Papers in Metaphysics and Epistemology*. Cambridge University Press, Cambridge, 1999.

Alyssa Ney. Physicalism and our knowledge of intrinsic properties. *Australasian Journal of Philosophy 85 (1)*, pages 41–60, 2007.

Stathis Psillos. Ramsey's *Ramsey-sentences*. In Maria Carla Galavotti, editor, *Cambridge and Vienna: Frank P. Ramsey and the Vienna Circle*, volume 12 of *Vienna Circle Institute Yearbook*, pages 67–90. Springer, 2006.

Hilary Putnam. Models and Reality. *Journal of Symbolic Logic 45*, pages 464–482, 1980.

Hilary Putnam. *Reason, Truth, and History*. CUP, 1981.

Jonathan Schaffer. Quiddistic Knowledge. *Philosophical Studies*, 23:1–32, 2005. Reprinted in Jackson and Priest 2004.

Jonathan Schaffer. Knowing the answer. *Philosophy and Phenomenological Research*, 75, 2007.

Robert C. Stalnaker. *Ways a World Might Be. Metaphysical and Anti-Metaphysical Essays*. OUP, 2003.

Robert C. Stalnaker. Varieties of supervenience. *Philosophical Perspectives 10*, pages 221–241, 1996. Reprinted, with new appendices, in Stalnaker 2003, pp. 86-108.

Patrick Suppes. *Introduction to Logic*. Van Nostrand, 1957.

Ann Whittle. On an argument for humility. *Philosophical Studies*, 130:461–497, 2006.

Dean Zimmerman. One really big liquid sphere: Reply to Lewis. *Australasian Journal of Philosophy*, 77:213–215, 1999.