HUMANAMENTE

ISSUE 19 - DECEMBER 2011

Composition, Counterfactuals and Causation

EDITED BY
Massimiliano Carrara, Roberto Ciuni, Giorgio Lando

Edizioni ETS

Editorial Board

Editorial Staff

TABLE OF CONTENTS

BOOK REVIEWS

REPORTS

Introduction
Composition, Counterfactuals, Causation

The problems of how the world is made, how things could have gone, and how causal relations work (if any such relation is at play) cross the entire historical development of philosophy. In the last forty years, the philosophical debate has given these problems a prominent role in its agenda, and David Lewis has suggested methodologies and theories that have contributed to enrich our notions in the fields of mereology, modality and the theory of causation. Such contributions have been among the most influential in analytic philosophy. The following theses – among others – have been milestones for the current philosophical debate:

1) The relation of parthood is exhaustively characterized by classical extensional mereology.
2) Every contingent feature of our world is determined by some "fundamental" physical properties and by the spatio-temporal relations between their bearers.
3) The evaluation of any counterfactual conditional involves a particular kind of similarity relation between the actual world and the scenario presented by the conditional.
4) Causation consists in a particular form of counterfactual link between the cause-event and the effect-event.

Many different philosophers have taken the above issues as points of reference, while revising in depth important aspects of the general picture. Some philosophers have accepted the idea of a mereological constitution of the world, while at the same time admitting universals as relata of the part-whole relation – contrary to Lewis's nominalist stance. Many have accepted that counterfactuals are to be explained as a linguistic and modal phenomenon - not as a meta-linguistic one – but dismiss any analysis of counterfactuals in terms of similarity. Many accept the idea that causality implies some "resistance to counterfactual situations", and yet find the details or methodology of Lewis's theory of causation quite unsatisfactory.

The views evolving from the criticisms or acceptance of 1-4 are still lively and ongoing. The idea behind this issue is to offer a ready-made representation of the most recent theories and position which are emerging in the debate and take David Lewis as their main theoretical source, critical target, or point of departure.

A possible point of originality in our approach is that the most famous Lewis's doctrine – the so-called modal realism, according to which possible worlds are concrete entities – is not given prominent attention. Modal realism is still a lively option in the philosophical debate and has been important for Lewis also as a source of ideas about seemingly unrelated topics. However, as Lewis himself was ready to admit, modal realism is also a highly controversial and counter-intuitive doctrine, and often elicits harsh reactions. Since many (and perhaps most) Lewis's contributions to the theory of constitution, counterfactuals and causation can be adopted or rejected independently of modal realism, it would be unwise to be driven primarily by these harsh reaction when dealing with these topics. For this reason, our approach is to keep modal realism as apart as possible.

The first part of the issue includes six original papers. Some of them share a common theme, and the order tries to respect these affinities. In the first paper ("Counterfactual Fallacies"), Andrea Iacona argues, *contra* Lewis, that counterfactuals can be formalized as strict conditionals (that is as conditionals prefixed with a modal operator), without incurring in fallacious inferences. The proposed formalization incorporates Lewis's analysis in terms of similarity in the content of the strict conditional.

In his "The Large-Scale Joints of the World", Ned Hall takes Lewis's thesis that there are objective joints in nature as his starting point, while admitting that *(i)* there is no straightforward way to specify how lower-level objective joints determine or ground the objective joints which should characterize the higher levels; *(ii)* Lewis's notion of causal explanation – needed for addressing point *(i)* – does not come with the required explanatory depth. In order to block these two problems, Hall proposes that causal explanations must provide causal information in a cognitively effective way, and admits that the naturalness of joints may come in *degrees*. The resulting perspective is quite interesting: though there is a fundamental level of the world whose features do not depend on our cognitive enterprise, there is *no* single way to describe its causal structure when it comes to higher levels (at least). By imposing some

taxonomies, we select the description of the causal structure that prove cognitively more effective.

The next two papers concern the theory of constitution, and are in general quite critical with Lewis's conviction that classical extensional mereology is the general, exhaustive theory of constitution. In "Any Sum of Parts of Water is Water", Henry Laycock provides some reasons to doubt the common assumption that the semantics of mass-terms (such as "water") directly satisfy mereological principles. Laycock assumes that mereology is an unsatisfying theory of constitution for structured organisms or artifacts, and aims to dispel also the illusion that the references of mass-terms are instead examples of concrete entities for which, in particular, extensionalism and unrestricted composition hold true.

Massimiliano Carrara and Enrico Martino ("Four Theses on the Alleged Innocence of Mereology") choose as critical target the idea that mereology is "ontologically innocent", in the sense that, given certain objects, there is no further ontological commitment to their mereological sum. Their conclusion is that the arguments for the innocence thesis are not conclusive, and that the thesis itself is ambiguous. Moreover, they construct a mereological model for a substantive fragment of set theory, arguing against the innocence of mereology with a strategy already adopted by Quine for second-order logic.

Andrea Borghini and Giorgio Lando deal with a controversial aspect of Lewis's doctrine of natural properties. According to a common interpretation of this doctrine, natural or fundamental properties, upon which every contingent feature of the world is expected to supervene, are instantiated only by minimal physical particles. In "Natural Properties, Supervenience, and Mereology", they argue that there are reasons internal to a broadly Lewisian kind of metaphysics (and in particular to mereology) to think instead that natural properties are spread at several levels of reality.

In the last paper, Giuliano Torrengo ("The Modal Dimension") discusses the hypothesis that modality is a genuine, fifth dimension of entities, on a par with spatial dimensions and time. His analysis focuses in particular on the so-called *tensed properties* and on the ontological status of modal wholes.

The issue includes also a large section of book commentaries. First of all, Lewis's books which are relevant for our topics are analysed, in order to see what in them is still relevant and worth debating and what – on the contrary – can be considered obsolete. Vittorio Morato delves into the many philosophical issues connected to *Counterfactuals*, compares Lewis's

semantics for counterfactuals with the then dominant approaches, presents the semantics devised by Lewis and comments on some interesting – and sometimes overlooked – conceptual issues which arise from such a semantics. In particular, Morato comments on the failure of *Conditional Excluded Middle* and some other laws, which set apart Lewis's theory of counterfactuals from Stalnaker's, and stresses an interesting fact: some problems in deontic logic and in the logic of counterfactuals seems to be solved by employing basically the same strategy. Louis deRosset reviews the most important points of Lewis's masterpiece *On the Plurality of Worlds* and some objections raised against them in the literature, concluding that the continuing influence of this book on contemporary philosophy depends more on its methodological assumptions (in particular the kind of reductionism it proposes) than on the specific contents of modal realism. Finally, Einar Bohn chooses to concentrate on a single topic of *Parts of Classes*: the so-called doctrine of "composition as identity". According to Bohn, Lewis's version of this doctrine is that composition is a genuine case of identity, and not merely that composition is analogous to identity under certain respects, as it has been often interpreted. Once the thesis is properly understood in this way, it is possible to look more closely at its problems and hidden assumptions.

After that, two commentaries are devoted to books in philosophy of physics, where some pivotal theses of Lewis's metaphysics are criticized from the point of view of contemporary physics. George Darby discusses *What's Wrong with Microphysicalism?* by Hüttemann, while Emiliano Boccardi deals with *The Metaphysics within Physics* by Maudlin. These books are representative of a quite wide-spread critical attitude towards the prevailing *a priori* character of metaphysics in the Lewisian tradition and in favour of a more scientifically informed approach. Darby in particular, while being sympathetic with this kind of scientism, shows that the Lewisian metaphysician could still reply in an interesting way to some of the objections and that, on the other side, also the theses more directly inspired to contemporary science are sometimes in need of a philosophical, *a priori* clarification.

Laura Castelli and Tuomas Tahko comment upon two books where Lewis's mereological theory of ontological constitution is compared and contrasted with ancient philosophy. These books are interesting for two main reasons: first, Lewis's definite and radical stance on these issues comes out as a useful point of reference and conceptual tool also for historians of philosophy; second, they show that classical extensional mereology, assumed by Lewis to

be an innocent, crystal-clear and general theory, is in sharp contrast with some important philosophical traditions, in particular when counter-intuitive principles such as unrestricted composition and composition as identity are at stake. Castelli analyses *Plato on Parts and Wholes* by Harte, where Plato's intensional approach to parthood is reconstructed and deemed to be preferable to Lewis's extensionalism, while Tahko scrutinizes Koslicki's development of an Aristotelian hylomorphic theory of constitution in her *The Structure of Objects*.

The last commentary, by Daniele Chiffi and Silvia Gaio, is about an important paper by Sider, "Temporal Parts". Sider's views in this paper about mereology and philosophy of time are deeply influenced by Lewis.

The issue is completed by a review and a conference report. Giulia Felappi reviews *Metametaphysics*, edited by Chalmers, Manley and Wassermann. This collection of papers is devoted to understanding whether metaphysics – Lewis's one included – is possible, useful and what it is really about. Finally, Adriano Angelucci gives an overview of a conference on the philosophy of Lewis held in Urbino in June 2011, another important sign that the interest in his thought is alive and widespread.

Massimiliano Carrara
massimiliano.carrara@unipd.it
Associate Professor, Department of Philosophy, University of Padua

Roberto Ciuni
roberto.ciuni@rub.de
Humboldt Postdoctoral Fellow , Institut für Philosophie II, Ruhr-Univesität Bochum

Giorgio Lando
giorgio.lando@sns.it
Postdoctoral Fellow, Department of Humanities, Scuola Normale Superiore

Counterfactual Fallacies *

Andrea Iacona†
ai@cc.univaq.it

ABSTRACT

A widely accepted claim about counterfactuals is that they differ from strict conditionals, that is, there is no adequate representation of them as sentences of the form $\Box\alpha \supset \beta$. To justify this claim, Stalnaker and Lewis have argued that some fallacious inferences would turn out valid if counterfactuals were so represented. However, their argument has a flaw, as it rests on a questionable assumption about the relation between surface grammar and logical form. Without that assumption, no consequence of the alleged kind is obtained, hence the claim may be rejected.

1.

A counterfactual is a conditional 'If it were the case that *p*, then it would be the case that *q*', where '*p*' is the antecedent and '*q*' is the consequent. For example, the following sentence is a counterfactual:

(1) If kangaroos had no tails, they would topple over

The obvious paraphrase of (1) is 'If it were the case that kangaroos have no tails, then it would be the case that they topple over'. A strict conditional is a sentence of the form $\Box\alpha \supset \beta$. In the familiar semantics of modal logic, $\Box\alpha \supset \beta$ is true in a world *w* if and only if $\alpha \supset \beta$ is true in every world accessible from *w*. If we call α-world a world in which α is true, this means that $\Box\alpha \supset \beta$ is true in *w* if and only if β is true in every accessible α-world. So it is tempting to say

* Many thanks to Andrea Borghini and José Diez for their comments on previous versions of this paper.
† University of L'Aquila, Italy.

that a counterfactual that has '*p*' as antecedent and '*q*' as consequent – a *p*/*q* counterfactual from now on – is a strict conditional that is true if and only if '*q*' is true in every world of some suitably restricted set in which '*p*' is true.[1]

However, Stalnaker and Lewis have argued that this temptation must be resisted. A strict conditional analysis of counterfactuals may appear tenable when one looks at this or that counterfactual, but it proves inadequate if one reflects on sets of counterfactuals and the logical relations they involve. At least three basic inference rules that hold for strict conditionals do not hold for counterfactuals, that is, there are at least three distinctive "counterfactual fallacies". The first is the fallacy of *strengthening the antecedent*. Consider the argument A1:

(2) If Otto had come, it would have been a lively party

---

(3) If Otto and Anna had come, it would have been a lively party

Imagine that Otto is a cheerful person, but that he just broke up with Anna after six months of endless rows. In such a situation (2) may be true even though (3) is false. In other words, (2) is consistent with

(4) If Otto and Anna had come, it would have been a dreary party

Therefore, A1 is invalid. But the following schema, S1, is valid:

$$\frac{\Box\alpha \supset \beta}{\Box (\alpha \wedge \gamma) \supset \beta}$$

For if β is true in all accessible α-worlds, *a fortiori* it will be true in all accessible α-worlds in which γ is true. This means that A1 cannot be represented as an instance of S1.[2]

The second is the fallacy of *transitivity*. Consider the argument A2:

[1] Mayo (1957) is among the early works in which it is suggested that counterfactuals amount to strict conditionals.

[2] Stalnaker (1991, p. 38); Lewis (1973, pp. 10–13 and 31). The sequence formed by (2) and (4) is called a Sobel sequence, from Lewis(1973, p. 10, fn).

(5) If Anna had gone to the party, Waldo would have gone
(6) If Otto had gone to the party, Anna would have gone

---

(7) If Otto had gone to the party, Waldo would have gone

Imagine that Waldo fancies Anna, although he never runs the risk of meeting his successful rival Otto. Imagine also that Otto was locked up at the time of the party, so that his going to the party is a remote possibility, but that Anna almost did go, as she hoped to meet him. In such a situation (5) and (6) may be true even though (7) is false. Therefore, A2 is invalid. However, the following schema, S2, is valid:

$\Box\beta \supset \gamma$
$\Box\alpha \supset \beta$

$\Box\alpha \supset \gamma$

For if all accessible α-worlds are β-worlds and all accessible β-worlds are γ-worlds, then all accessible α-worlds are γ-worlds. So A2 cannot be represented as an instance of S2.[3]

The third is the fallacy of contraposition. Consider the argument A3:

(8) If Otto had gone to the party, Anna would have gone

---

(9) If Anna had not gone, Otto would not have gone

Imagine that Otto wanted to go to the party but stayed away just to avoid Anna, while Anna would definitely have gone if Otto had been around. In such a situation (8) may be true even though (9) is false. Therefore, A3 is invalid. However, the following schema, S3, is valid:

[3] (Stalnaker, 1991, p. 38; Lewis, 1973, pp. 32–33). Note that S2 entails S1, as it is easily seen if α is replaced with $\alpha \wedge \beta$. So the failure of S1 alone suffices to discard S2.

$$\frac{\Box\alpha \supset \beta}{\Box\neg\beta \supset \neg\alpha}$$

For $\alpha \supset \beta$ and $\neg\beta \supset \neg\alpha$ have the same truth-value in every world. This means that A3 cannot be represented as an instance of S3 (Lewis, 1973, p. 35; Stalnaker, 1991, p. 39).

The Stalnaker-Lewis argument may be summarized as follows. If counterfactuals are strict conditionals, then A1–A3 instantiate S1–S3. But that is absurd. A1–A3 are invalid arguments, while S1–S3 are valid schemas. So counterfactuals are not strict conditionals. This paper is intended to provide a reason to doubt the Stalnaker-Lewis argument.

2.

The line of resistance that will be suggested differs from at least three objections that may be prompted by some contextualist accounts of counterfactuals as strict conditionals that have emerged recently. The assumption that the three objections share is that counterfactuals are highly context-sensitive strict conditionals, in that the accessibility relation associated with them varies as a function of their antecedent. On this assumption, counterfactuals with different antecedents are intuitively assessed relative to different contexts, because their antecedents select different sets of relevantly similar worlds.[4]

The first objection goes as follows. It is wrong to assume that A1–A3 are invalid arguments. In order to evaluate A1–A3, just as any other argument affected by context-sensitivity, the context must be held fixed. An argument is valid if and only if, for every context, if the premises are true relative to that context then the conclusion is true relative to that context. But A1–A3 are such that there is no context relative to which the premises are true and the

[4] The supposition that the counterfactuals in a Sobel sequence – hence in A1 – are strict conditionals that involve different contexts, initially dismissed in (Lewis, 1973, p. 13), is developed in (von Fintel, 2001) and in (Gillies, 2007). (Lowe, 1995, pp. 56–57), suggests that arguments such as A2 can be treated as cases of equivocation due to context-sensitivity.

conclusion false, hence they are valid. The invalidity of A1–A3 is only apparent, due to the context-shifts in their intended reading.[5]

This objection is not entirely convincing. Even if one grants that the counterfactuals in A1–A3 involve different contexts, and that no context makes the premises true and the conclusion false, one is not compelled to conclude that A1–A3 are valid. Certainly, the definition of validity as truth-preservation in any context entails that conclusion, so it clashes with our inclination to regard A1–A3 as invalid. But this clash does not show that our inclination is misplaced more than it shows that the definition is unable to handle such cases. In what follows it will be taken for granted that A1–A3 are invalid, just as they appear.

The second objection is opposite to the first, as it attacks the assumption that S1–S3 are valid schemas. A proponent of the view that counterfactuals are highly context-sensitive strict conditionals may grant that A1–A3 are invalid arguments and that A1–A3 instantiate S1–S3, but claim that S1–S3 are invalid precisely in virtue of that fact. For a schema is valid just in case all its instances are valid arguments.

This objection throws the baby out with the bathwater. To deny that S1–S3 are valid schemas is to deny the basic principles of modal logic. For the validity of S1–S3 follows from those principles. If S1–S3 are invalid, then the semantics of the language in which they are expressed is not the familiar semantics of modal logic, and □ does not have its familiar meaning. Even if one is willing to accept this consequence, which is not easy to swallow, the question remains of how one can maintain the claim that counterfactuals are strict conditionals in some sense that is relevant to the Stalnaker-Lewis argument. For that argument is intended to dismiss the claim that counterfactuals are strict conditionals in the familiar sense.

The third objection goes as follows. A1–A3 are invalid arguments, S1–S3 are valid schemas, but there is nothing absurd in the supposition that A1–A3 instantiate S1–S3. When □ occurs more than once in an argument and it is associated with different accessibility relations, the possibility that the premises of the argument are true and the conclusion false is not detectable

[5] A reasoning along these lines is offered in Brogaard & Salerno (2008), although it is not accompanied by a strict conditional analysis of counterfactuals. Cross (2011) questions the contextualist assumptions that underlie that reasoning.

from its logical form. In other words, the invalidity of A1–A3 is not amenable to formal explanation.

This objection is defeatist in at least one important respect. As long as formalization is understood in the usual way as a representation of logical form that displays fundamental logical properties such as validity, it is hard to make sense of the claim that A1–A3 are invalid arguments that instantiate S1–S3. To say so is to say something odd, namely, that although it is correct to represent the counterfactuals in A1–A3 as strict conditionals, such representation plays no role in a formal explanation of the logical properties of A1–A3. Nothing like this will be suggested here. Logical form does play a role in formal explanation, hence the logical properties of A1–A3 must be detectable from the logical form of the counterfactuals in them.

3.

So far there is nothing to object to the Stalnaker-Lewis argument. A1–A3 are invalid arguments, S1–S3 are valid schemas, and the supposition that A1–A3 instantiate S1–S3 leads to absurdity. The flaw of the argument lies elsewhere, namely, in the assumption that if counterfactuals are strict conditionals then A1–A3 instantiate S1–S3. Presumably, the rationale for this assumption is that the only way to represent a *p*/*q* counterfactual as a strict conditional is to suppose that its logical form is expressed by a formula $\Box\alpha \supset \beta$ where α stands for '*p*' and β stands for '*q*'. But that is not the only way, nor is the best. There is another way to represent a *p*/*q* counterfactual as a strict conditional, which is in accordance with the plausible hypothesis that the meaning of the counterfactual is that in any possible world in which *p*, and which resembles our world as much as the supposition that p permits it to, *q*. The view is that the logical form of a *p*/*q* counterfactual is $\Box\alpha \supset \beta$, where α does not stand for '*p*' but for the stronger condition that *p and for the rest things are like in our world as much as the supposition that p permits it to*. For example, in the case of (1) α expresses the condition that kangaroos have no tails and for the rest things are like in our world as much as kangaroos having no tails permits it to. That is, if γ stands for 'Kangaroos have no tails', α amounts to $\gamma \wedge \delta$, where δ expresses the similarity constraint required. The idea that underlies this view turns out clear if one reflects on the contrast between a *p*/*q* counterfactual and an overt strict conditional 'Necessarily, if *p* then *q*'. Consider (1) and the following sentence:

(10) Necessarily, if kangaroos have no tails, then they topple over

While the truth condition of (10) is that kangaroos topple over in any possible world in which they have no tails, the truth-condition of (1) is that kangaroos topple over in any possible world such that kangaroos have no tails and things are like our world as much as the supposition that kangaroos have no tails permits it to. Now consider a formal representation of (1) as $\Box\alpha \supset \beta$. If the same formula were assigned to (10), as required by the supposition that α stands for 'Kangaroos have no tails', there would be no way to distinguish (1) from (10) by looking at its formal representation. But this would go against something that is usually taken for granted about formalization, namely, that sentences with different truth-conditions are to be represented by means of distinct formulas, that is, formulas that can have different truth-values in the same interpretation. It is natural to expect that the difference in truth-conditions between (1) and (10) is formally represented, so that the corresponding formulas have different truth-values in some interpretation. Or at least, this is what Stalnaker, Lewis and many others would say. The simplest way to draw the distinction is to assign a different formula $\Box\alpha \supset \beta$ to (10), assuming that γ stands for 'Kangaroos have no tails' while α amounts to a stronger condition $\gamma \wedge \delta$. In substance, the idea is that the logical form of counterfactuals systematically diverges from their surface grammar, in that the antecedent of the formula that expresses their truth-condition does not correspond to their antecedent. In that sense counterfactuals differ from overt strict conditionals, whose antecedent is stated explicitly.[6]

On this view, A1–A3 do not instantiate S1–S3. Consider A1. If (2) is represented as $\Box\alpha \supset \beta$, then α does not stand for 'Otto has come' but for 'Otto has come and for the rest things are like in our world as much as Otto coming permits it to'. So (3) cannot be represented as $\Box (\alpha \wedge \gamma) \supset \beta$. Rather, it is to be represented as $\Box\gamma \supset \beta$, where γ expresses a condition that entails 'Otto and Anna has come' but is not reducible to a conjunction that includes α. For one thing is to require that a world is similar to ours as much as the truth of 'Otto has come' permits it to, quite another thing is to require that a world is

[6] In a longer paper, *Counterfactuals as Strict Conditionals*, I spell out the view that counterfactuals are strict conditionals whose antecedent is stated elliptically, and compare it with the account of counterfactuals suggested by Stalnaker and Lewis.

similar to ours as much as the truth of 'Otto and Anna has come' permits it to. Therefore, the schema instantiated by A1 is not S1 but the following, S4:

$$\frac{\Box\alpha \supset \beta}{\Box\gamma \supset \beta}$$

Consider A2. If (6) is represented as $\Box\alpha \supset \beta$, then (5) cannot be represented as $\Box\beta \supset \gamma$ but rather as $\Box\gamma \supset \delta$, where γ entails β. Therefore, the schema instantiated by A2 is not S2 but the following, S5:

$$\frac{\begin{array}{l}\Box\gamma \supset \delta \\ \Box\alpha \supset \beta\end{array}}{\Box\alpha \supset \delta}$$

Finally, consider A3. If (8) is represented as $\Box\alpha \supset \beta$, the antecedent of the formula that represents (9) cannot be $\neg\beta$ but a different formula γ that entails $\neg\beta$. Similarly, its consequent cannot be $\neg\alpha$ but a different formula δ that stands for 'Otto has not gone'. Therefore, the schema instantiated by A3 is not S3 but the following, S6:

$$\frac{\Box\alpha \supset \beta}{\Box\gamma \supset \delta}$$

Since S4–S6 are invalid schemas, the invalidity of A1–A3 is easily explained. A fallacy is a bad argument that may appear good at first sight, and counterfactual fallacies are no exception in this respect. A1–A3 may seem valid, in that the antecedents of the counterfactuals they contain make them look similar to other arguments that instantiate valid schemas. But in reality they are invalid, since they do not instantiate those schemas.

REFERENCES

Brogaard, B., & Salerno, J. (2008). Counterfactuals and context. *Analysis*, *68*, 39–46.

Cross,C.B. (2011). Comparative world similarity and what is held fixed in counterfactuals. *Analysis*, *71*, 91–96.

Gillies, A.S. (2007). Counterfactual scorekeeping. *Linguistics and Philosophy*, *30*, 329–360.

Lewis, D. (1973). *Counterfactuals*. London: Blackwell.

Lowe, E.J. (1995). The truth about counterfactuals. *Philosophical Quarterly*, *45*, 41–59.

Mayo, B. (1957). Conditional statements. *Philosophical Review*, *66*, 291–303.

Stalnaker, R. (1991). A theory of conditionals. In F. Jackson (Ed.), *Conditionals*. Oxford: Oxford University Press, 28–45.

von Fintel, K. (2001). Counterfactuals in a dynamic context. In M. Kenstowicz (Ed.), *Ken Hale: A Life in Language*. Cambridge, MA: The MIT Press, 123–152.

The Large-Scale Joints of the World

Ned Hall †
ehall@fas.harvard.edu

1. Introduction

What is the compositional structure of reality?

That question divides naturally into these two: What is the compositional structure of the *particulars* that populate reality? And what is the structure of the *properties* and *relations* that fix what these entities are *like*?

David Lewis's work in ontology and mereology provides the materials for an extraordinarily clean answer to the first question. First, among the particulars[1] that populate reality are *mereological simples*: entities that have no proper parts. (A plausible candidate for these simples: spacetime points.) Second, every collection of such entities has a *unique mereological fusion*. And third, every particular is either a simple, or a fusion of simples.[2] That's it.

I propose to take this answer on board.[3] What, then, about our second question? Here it looks as though we can draw on an additional Lewisian thesis:

> **Joints**: There is a distinction – at the level of metaphysics – between *more and less natural properties*. Some properties (*having mass 1 gram*, perhaps) are *perfectly natural*; others (*being a methane molecule*, perhaps) are less-but-still-quite natural; still others (*being grue* is a favorite) are not very natural at all.

† Harvard University, MA, USA.

[1] Note that the restriction to "particulars" is in place because Lewis allows that there might be *other* sorts of entities – e.g., repeatable universals.

[2] I said that Lewis's work provides the materials for this answer. It is much less clear whether Lewis himself endorsed this answer; the textual evidence is somewhat ambiguous. At any rate, if he did, he endorsed it as a contingent thesis; see for example the opening section of his 1994. I myself prefer the view that this answer, if correct, is metaphysically necessary but a posteriori; see my 2011a.

[3] Though only for the purposes of this essay: there are, after all, reasonable grounds for reservations. What about *holes*, for example? The gyrations Argle goes through (Lewis & Lewis, 1970) to accommodate them suggest to me that a more relaxed view in ontology is called for, one that agrees that the existence of any non-fundamental particular must be appropriately *grounded* in facts about fundamental entities – viz., mereological simples – without agreeing that every non-fundamental particular must be *composed* of fundamental ones.

> This distinction earns its philosophical keep because of the number and centrality of the philosophical projects that must presuppose it. And to say that this distinction resides at the level of metaphysics is, at least in part, to say that it is not grounded in facts about human psychology.

What Joints tells us, in effect, is that there are *objective joints in nature* that appear at different mereological scales, constituted by the pattern of instantiation of natural properties and relations by the particulars that exist at those scales. That is an attractive picture, but please note that it is nothing more than a picture. Whereas we were able to get a complete and exact answer to the first of our two questions about the compositional structure of reality – thanks to the fact that the version of mereology that our answer drew upon was itself clear and exact – what we have here is no more than a framework for such an answer. It needs to be filled out in at least two ways: we need an account of what naturalness of properties and relations is; and we need an account of how naturalness at one scale fits together with naturalness at other scales.

Now, these accounts ought, I think, to be constrained by the need to make sense of one of the central aims of *empirical inquiry*, especially *mature* empirical inquiry – i.e., *scientific* inquiry. And that aim is to provide us inquirers with *explanations* for why our world behaves the way it does. As it happens, Lewis (1986b) also defended an interesting and important thesis about what such explanations consist in:

> **Causal Explanation**: To explain an event is to give some information about its history of causes. Since causation is both *transitive* – if event A is a cause of event B, and B of C, then A is thereby a cause of C – and *egalitarian* – even background 'enabling conditions' of an event count as causes of it, notwithstanding the oddity of saying so in ordinary conversation – the causal history of a typical event will almost certainly be vast beyond any possibility of full and accurate conveying. And so, in any particular context in which we seek understanding of why some event occurred, pragmatic factors will play a large role in fixing how much and which parts of that event's causal history ought to be highlighted. But that is a quite unexceptional intrusion of pragmatics, and one that ought to make no difference to the philosophical project of saying what sort of information explanatory information *is*. That project is completed – in the case of events, anyway – once we identify information explanatory of them with information about their causal histories.

What I would like to explore in this essay are the prospects for fleshing out Joints and Causal Explanation in a way that makes for a unified package.

Troubles will quickly appear: as to Joints, it will emerge that there is no obvious way to say how the distinctions in naturalness that reside at the most fundamental mereological scales ground such distinctions at higher scales, while preserving the view that these latter distinctions are (like the former) perfectly *objective*. As to Causal Explanation, it will emerge that Lewis has overlooked something of great importance to us as inquirers who seek to understand our world, which is knowledge not merely of the *causes* of some given phenomenon, but knowledge of that in virtue of which the causes *are* causes. Only when augmented by this latter sort of information does a causal explanation of some event achieve the right sort of explanatory *depth* (to borrow Michael Strevens's apt expression; see (Strevens, 2009).)

The troubles for Joints and the troubles for Causal Explanation are, I think, connected. To say exactly why will require a bit of spelling out, so here I will just offer some teasers.

To achieve a decent philosophical account of what explanatory *depth* consists in, we will need to supplement Lewis's idea that explanatory information is causal information with the distinctively unificationist idea that we improve our understanding of our world by finding ways to *organize our information about it in a cognitively effective fashion*. Adding this dose of unificationism yields a view according to which, at least at scales above the most mereologically fundamental, the distinctions that we as empirical inquirers find most explanatorily valuable to draw *derive* their explanatory value in part from the way in which they collectively organize, in a cognitively effective manner, our view of the subject matter that is the target of our inquiry. So it turns out that it is partly a matter of human psychology what makes for an explanatorily valuable distinction (at least, on the assumption that it is in part a matter of human psychology what makes for effective cognitive organization). So, since talk of nature's "joints" *just is* talk of those distinctions in nature grasp of which is essential for explanation, it follows that the joints in nature (at least, at scales above the most mereologically fundamental) are what they are in part because of facts about human psychology.

But lest you think I'm succumbing to the "postmodern forces of darkness" (to use Sider's delightful phrase; see his 2011), let me highlight two important qualifications: first, nothing in what I will argue will suggest that the most *fundamental* joints in nature – the joints that it is the job of fundamental physics to discern – are to any extent of human origin. Second, the conception of nature's joints that I will sketch is perfectly consistent with the view that the

world possesses a rich and completely objective causal structure. It is just that in organizing our view of this causal structure, we must impose on it certain taxonomies. And the way in which we do so – more exactly, the fact that certain ways of imposing taxonomies are *explanatorily better* than others – cannot itself be explained merely by reference to that causal structure: that structure does not, as it were, force upon us, merely by virtue of its internal nature, certain ways of organizing it.

Let me now try to unpack all this. We'll begin by digging a little more deeply into Joints.

2. Natural, non-natural, more and less natural

Let's start by getting a little clearer on what the natural/non-natural distinction *is*, and then reviewing Lewis's case for taking the distinction on board. After that, we can consider the complications that arise from trying to give a philosophical account of how this distinction can come in *degrees*.

The distinction is meant to divide properties and relations into those that are somehow *genuine* – reflecting or constitutive of real distinctions in the world – and those that are somehow *artificial* or *gerrymandered* – reflecting, perhaps, nothing more than an arbitrary, purely conventional decision to use a certain label in a certain way. Here is another way to put the idea. Consider all the entities that there are, or that there possibly could have been; indulge, for the moment, Lewis's odd view that the latter sorts of entities exist in exactly the same sense as the former. (That indulgence will make things simpler; it's not essential.) Consider all the *sets* that can be formed from these elements – the power set of reality, if you like. Some of these sets will group together entities that, somehow, *belong* together: all the actual and possible electrons, say, or all the methane molecules, or all the wombats. Others will fall short of this ideal. There is the set that contains all the electrons, *and* all the wombats. Or the set that contains all the electrons that exist in worlds with at least one wombat, together with all the protons that exist in worlds with no wombats. And so on. Once you see what's going on, you'll see that the *vast majority* of these sets will fail to group together entities that are alike in some genuine respect (and fail much more dramatically than the two foregoing examples – both of which, after all, could at least be described in English). For short: among all the groupings

that, set-theoretically speaking, there are, only a small minority correspond to *real* distinctions in the world.[4]

Why believe in such a distinction among distinctions? Well, it's likely that you already *do* – at least, if you understood the brief exposition given in the last paragraph. But for all that, you might reasonably doubt that the natural/non-natural distinction is *objective*, in the sense that *what it is* for a property or relation to count as "natural" has nothing to do with human psychological responses to the world we inhabit. To begin to rebut this worry, as well as to flesh out our explication of the distinction, we should appreciate two reasons to endorse it that draw on aspects of our ordinary thought and talk about the world that are so intimate and familiar as to readily escape notice.

The first reason has to do with *change*. When Billy falls in love with Suzy, that is a genuine change in Billy; but it is not a genuine change in Suzy, notwithstanding the fact that she goes from lacking the property of being loved by Billy to having this property. So – and this is, of course, a perfectly familiar point in the philosophical literature – not every gain or loss of a property by a thing counts as a genuine or objective change in that thing. But if we maintain (as we should) the idea that every change in a thing *is* a gain or loss of a property by that thing, and that at least some change is a perfectly objective feature of the world, then we need an objective distinction among properties to say which gains or losses of properties count.

Second, while some similarities and differences among entities are no doubt in the eye of the beholder, some are not. Two methane molecules are more similar to each other than either is to a tomato, period. A comprehensive scheme for taxonomizing the items that populate our world that failed to recognize this fact would, whatever its other virtues, fall short in one epistemically crucial respect: it would fail to correctly limn one aspect of the world's structure. If we take this sort of structure to be an objective feature of the world – again, as it seems we should – and we take it to be constituted by the facts about which properties entities share or fail to share, then we need an objective distinction among properties to say which are those whose pattern of

[4] The last few sentences have tacitly restricted our attention to *properties*. But the natural/non-natural distinction applies to *relations*, as well. To handle, say, two-place relations, we should start by considering all possible *pairings* of (actual and possible) entities, and then consider all the sets of those pairings, the vast majority of which will correspond, so the thought goes, to no genuine way in which two things can be related. And so on, for 3- and more place relations.

instantiation fixes the structure of the world. (Compare (Sider, 2011).) (Arguably, the first point, about change, is just an instance of the second: from a suitably exalted space-time perspective, to talk about change is just to highlight temporal aspects of the world's overall spatiotemporal structure.)

If we begin with this intuitive distinction between natural and non-natural properties and relations, we can get part way to Lewis's conception of *perfectly* natural properties and relations simply by maxing out one criterion: whereas sharing of natural properties makes for similarity, sharing of *perfectly* natural properties makes for *perfect* similarity, or *duplication*.

Consider the property of being a methane molecule. Any two things that share this property – i.e., any two methane molecules – will ipso facto be quite similar to one another. But they need not be perfect qualitative duplicates: their internal configurations might differ slightly, or they might be slightly different in composition (say, one contains a carbon-12 atom, where the other contains a carbon-13 atom). By contrast, sharing of perfectly natural properties is supposed to yield perfect qualitative similarity. Thus, Lewis (1983b, p. 27): «Two things are qualitative duplicates if they have exactly the same perfectly natural properties».[5]

But the way in which sharing of perfectly natural properties grounds facts about similarity goes beyond the requirement that two objects that instantiate exactly the same ones are perfect qualitative duplicates. Consider two Newtonian point-particles that are perfect duplicates, having, say, exactly the same values for mass and charge. Suppose that, in some appropriate units, each particle has mass 1 and charge 1. Now, *having mass 1* and *having charge 1*, let us agree, are examples of perfectly natural properties. By contrast, here are some of the properties the particles instantiate that fall short of being perfectly natural (on the grounds that the sharing of them does not count as a way of being genuinely similar): *having mass 1 or charge 2*, *having mass 2 or charge 1*, *having mass 1 or charge 1*. But, on the assumption that it is metaphysically impossible for a particle to have two distinct values of mass or charge, the

[5] It's clear in the context that Lewis intends the "if" to be understood as "iff". Also, a more careful formulation of the idea would be the following: Two things A and B are qualitative duplicates iff there is a one-one mapping between those parts of A (including A itself) that instantiate perfectly natural properties or relations and those parts of B that do so, such that whenever some part or parts of A (respectively, B) instantiate some perfectly natural property or relation, the corresponding parts of B (respectively, A) instantiate the very same property or relation.

sharing of these three properties guarantees perfect duplication (at least, with respect to mass and charge). In short: if we take the *complete intrinsic nature* of a thing to be what is shared between it and its perfectly duplicates, then the notion of "perfect naturalness" imposes an additional structure on these complete intrinsic natures. Thus, the intrinsic nature of one of our particles is constituted by its being a point particle, having mass 1, and having charge 1; not (e.g.) by its being a point particle, having mass 1 or charge 2, having mass 2 or charge 1, and having mass 1 or charge 1.

Furthermore, the perfectly natural properties are supposed to collectively constitute a kind of minimal supervenience basis for all of reality: that is, the *whole truth* about the qualitative structure of the world is supposed to be grounded in the pattern of instantiation of perfectly natural properties and relations, in a way that involves no redundancy. This requirement provides an extra reason for excluding such "disjunctive" properties as *having mass 1 or charge 2* from the ranks of the perfectly natural, as well as a reason for so excluding "conjunctive" properties such as *having mass 1 and charge 2* (which could have gained membership, if our sole criterion concerned whether sharing the property is a way for two things to achieve a sufficiently high degree of similarity).

A few observations about Lewis's distinction will be helpful for what follows. First, the philosophical importance of the distinction is not limited to its uses in analytic metaphysics (or the theory of reference – more on this, in a moment). It appears, in addition, to be crucial for articulating what is arguably one of the central aims of *physics*, which is to provide an inventory of the fundamental physical magnitudes of our world. For while Lewis takes it to be a job for philosophy to defend the claim that some properties and relations are perfectly natural, he rightly takes it to be a job for physics to figure out which perfectly natural properties and relations happen to characterize our world. Second, notice that as soon as we grant that physics *has* this job, we can see that the distinction we want is not, or at any rate should not be limited to, a distinction among properties and relations: at least for the purposes of physics, it should be seen as a distinction among *determinable magnitudes*. (For more on this point, see (Hall, 2010).) Third, it is an extremely plausible thesis – but not, I think, an indisputable one, given that the thesis is also, ultimately, *empirical* – that perfect naturalness of properties, relations, and magnitudes is closely connected to mereology, via the thesis that the only entities that genuinely instantiate perfectly natural properties, relations, or magnitudes are

mereological simples (perhaps, just points of space and time). There may be exceptions – for example, it may be that the topological structure of space-time is best understood as a perfectly natural feature of *it*, not reducible to perfectly natural properties and relations instantiated by its ultimate parts. But for purposes of this essay I will simply bracket this issue, and take for granted that as soon as we ascend to any mereological scale above that of fundamental physics, the sorts of properties, relations, and magnitudes we encounter cannot qualify as perfectly natural. (For clutter-reduction, I will also henceforth mostly speak just of "properties", even when relations and magnitudes are also intended).

Lewis offers a number of reasons, many persuasive and all intriguing, for thinking that we need to accept, as a fundamental metaphysical distinction, a distinction between perfectly natural properties and the rest. Set these reasons aside (you can find most of them in his 1983b). The question I wish to focus on, instead, is this: should we also take this distinction to be *graded*? That is, should we insist that, among all the properties that are *not* perfectly natural, some are nevertheless *more* natural than others? And if we do so insist, what sort of account can we give of what these gradations consist in?

Now, as to the first question, I'm going to assume that the answer is "yes". Lewis himself offers one very important reason that I'm going to set aside, which is that without such a graded distinction, it will be impossible to give a naturalistically acceptable account of the content of language and thought that does not face an insurmountable underdetermination problem (see the last section of his 1983b, as well as his 1984). I am, instead, go to lean upon a much more prosaic observation about scientific inquiry, which is just that, regardless of the mereological scale at which it operates, it seems to be a central and nonnegotiable feature of such inquiry that it aims to develop the right sorts of descriptive resources for describing the structure of the world at the given scale. There are, for example, ever so many ways that, logically speaking, one could describe reality at the scale at which chemistry operates. But the chemist's taxonomy is the best (or at the very least, one of the best); and I'm going to assume that to say that it is best is to say that this taxonomy tracks highly though not perfectly natural distinctions in nature.

So let us grant that we cannot do without a graded distinction between more and less natural properties. Then how do we answer the second question – how do we give an account of what these gradations consist in? That turns out, I think, to be an extremely difficult (and open) problem. For now, I would

like merely to consider and reject two approaches to it, as a way of highlighting how difficult it is.

First, one might simply treat gradations of naturalness as metaphysically primitive, and so unanalyzable. Superficially, this might seem an acceptable option; after all, Lewis himself argues that it is perfectly reasonable to accept the distinction between *perfectly* natural properties and the rest as primitive. But in fact I think this option is not acceptable. Here, briefly, are two reasons.

To begin with, we have granted that the pattern of instantiation of perfectly natural properties completely determines the qualitative structure of the world. And so, facts about comparative similarities and differences among objects that exist at scales above the most mereologically fundamental are fixed by this pattern of instantiation. But when, in the course of scientific (or even ordinary-life) investigation, we introduce distinctions among the less than perfectly natural properties, these distinctions earn their keep only insofar as they track explanatorily important similarities and differences among large-scale objects. So, if these explanatory distinctions are themselves ultimately grounded in facts about how the perfectly natural properties array themselves, then that just is to say that gradations in naturalness must be so grounded – and so cannot be metaphysically primitive, after all.

In addition, it's a good piece of philosophical methodology to avoid primitives that are ungainly. And in the present case, the imagined primitive seems *too* ungainly, since the way that gradations in naturalness are marked is quite complex. For example, for a chemist, the classification "being a methane molecule" will be more natural than the classification "being a methane molecule whose carbon atom is C-12": for whether the carbon atom is C-12 or C-13 will make no relevant *chemical* difference. But for a *nuclear* chemist, that difference *is* relevant, in a way that could reasonably reverse the judgment of naturalness. (It matters, for example, to the longevity of the methane molecule.) Cases like this suggest that the sort of naturalness that comes in degrees will exhibit a relativity to explanatory context: what count as the distinctions relevant to *some* explanatory projects may (even at the same mereological scale) differ from what count as the distinctions relevant to *distinct* explanatory projects. That complexity of conceptual structure makes it implausible that degrees of naturalness are simply metaphysically primitive, and at any rate deprives a philosophical account that treats them so of the resources it would need to explain this relativity to explanatory context.

So I'm going to henceforth assume that however attractive it may be to treat the category of *perfectly* natural property as primitive (in part, presumably, because the sort of relativity to explanatory context just discussed does not show up at the level of fundamental physics), an informative account is obligatory of the gradations in naturalness that the less than perfectly natural properties exhibit.

Lewis himself is perfectly aware of the need for an account of what makes one property *more natural* than another. He offers a simple and straightforward proposal, in the context of explaining how distinctions of naturalness yield distinctions in eligibility of reference:

> Indeed, physics discovers which things and classes are the most elite of all; but others are elite also, though to a lesser degree. The less elite are so because they are connected to the most elite by chains of definability. Long chains, by the time we reach the moderately elite classes of cats and pencils and puddles; but the chains required to reach the utterly ineligible would be far longer still. (Lewis 1984, p. 228)

Begin with a language whose non-logical vocabulary refers to *perfectly* natural properties; take this language to be rich enough that every perfectly natural property instantiated in our world gets referred to. Given some less-than-perfectly-natural property F, there will be *some* predicate of our canonical language – perhaps a very long, complicated predicate – that expresses it. F will be *more natural than* some other less-than-perfectly natural property G just in case the predicate expressing F is shorter than the predicate expressing G.

The proposal pretty clearly needs some refinement. *Many* predicates will express a given property; presumably we are to pick the shortest. While we can compare two predicates for length, we can also compare them for *simplicity*. Suppose the shortest predicate for F is slightly longer – but significantly simpler – than the shortest predicate for G;[6] should the advantage in simplicity outweigh the disadvantage in length? How shall we handle properties that are "multiply realizable", in the specific sense that predicates expressing them cannot be defined in a canonical language whose non-logical vocabulary refers only to *actual* perfectly natural properties? (For example, suppose the

[6] How would this happen? Well (for example), suppose the predicates contain quantifiers, and that those that appear in the predicate for F involve fewer alternations than those that appear in the predicate for G.

restriction of F to this world can be defined simply and efficiently, whereas F itself cannot; should we count F as natural-in-our-world?)

These are interesting questions, and perhaps even important, *if* our aim is to construct an account of the more/less natural distinction that will serve the purposes of a theory of the content of language and thought. But our aim here is different: it is to see how nature's ultimate joints – the joints given by the pattern of instantiation of perfectly natural properties, the joints it is the business of fundamental physics to discern – give rise to the less-than-ultimate joints whose study is the province of the special sciences. (Note well that I do *not* make Lewis's hasty assumption that a more/less natural distinction suited to *this* task will be one and the same as the more/less natural distinction suited to the theory of *content*.) And we do not need to resolve the foregoing questions to see that Lewis's proposal is on entirely the wrong track.

Consider the property of being methane. One early sign of trouble for Lewis's approach is that it looks as though, on that approach, it will be fixed once and for all whether this property is more natural than the property of being "stable" methane (that is, the property of being a methane molecule whose carbon atom is C-12), and that is just because the facts that Lewis's approach deems relevant – shortness and/or simplicity of definition in a canonical language – do not themselves exhibit any dependency upon explanatory context. There is, perhaps, the tiniest bit of wiggle room: maybe the trade-off between simplicity and length of definition could be taken to vary with explanatory context. But it would be pointless to pursue such a loophole, for the real, underlying trouble is much more straightforward: what this approach deems relevant to naturalness just bears no adequate connection to what underlies the explanatory utility of classifications in actual scientific practice.

A couple of examples should make the difficulty sufficient vivid. First, compare the property of *being methane* to the property of *being composed of 26 particles, each of which is a proton, neutron, or electron* (a property that most, but not all methane molecules share). Here is something indisputable: the shortest, simplest predicate in the canonical language expressing the first of these properties will be *vastly* longer and more complicated than the shortest, simplest predicate expressing the second. We *know* this: after all, I

just all but *gave* the canonical predicate expressing the second property[7], whereas, to produce the canonical predicate for the first property, one would need to begin with the standard chemist's definition of methane as "a molecule composed of four hydrogen atoms covalently bonded to a carbon atom", and proceed to unpack the predicates "is a hydrogen atom", "is a carbon atom", and "is covalently bonded to". And doing *that* will require bringing to bear substantial resources from theoretical chemistry and nuclear physics. It would take a while. You're rather unlikely to find anyone patient enough to be willing to pursue this project.

So, by the lights of Lewis's account of the less-than-perfectly-natural, the property of being composed of 26 protons, neutrons, and electrons ought to be *significantly* more natural than the property of being methane. And that's just silly. After all, methane molecules all have in common – in virtue of being methane molecules! – a wide variety of explanatorily important features. Thanks to the way that their covalent bonds affect their structure, they are all close to perfectly *tetrahedral*. They are all close to the same *size*. They all *react*, chemically, in exactly the same way. And so on. By contrast, *no explanatory purpose whatsoever* is served by distinguishing the class of things-composed-of-26-protons-neutrons-and-electrons. The vast majority of 26-pne's in the universe, after all, are *scattered*, their 26 different parts separated by light-years of space. And even those that are not exhibit no interesting or systematic behavior. (Except, of course, those that also happen to be methane molecules.) Just picture yourself writing a grant proposal – a *serious* one, mind you – asking for funding so that you can start a new program of research into 26-pne's.

Here is a second example, that reinforces the point that greater explanatory value of a classification can very often point in the *opposite direction* from greater simplicity or efficiency of canonical definition. Granted, *being methane* is a useful chemical property to know about. But *being a saturated hydrocarbon* is much *more* useful: it lends itself to a greater range of more important

[7] Assuming that protons, neutrons, and electrons are all fundamental particles, that is. Of course they're not (not protons and neutrons, anyway); but dropping this assumption would not make the slightest bit of difference to the plausibility of the claim that the canonical predicate expressing the second property will be vastly simpler and shorter than the canonical predicate expressing the first property.

generalizations, as any organic chemist will tell you.[8] So, measured by explanatory utility, the category of saturated hydrocarbons is a much more valuable one to distinguish than the category of methane molecules. But its canonical predicate will necessarily be longer and more complicated than the canonical methane-predicate. (Consider, for example, that this predicate must include a specification that the covalent bonds holding the molecule together are *single*, something that would be redundant in the methane-predicate.)

Examples like these convince me that there is something fundamentally misguided about Lewis's account of the less-than-perfectly-natural. It *may* be that brevity and/or simplicity of canonical definition plays some role in accounting for how nature's joint emerge at larger mereological scales. But it cannot be the whole story.

Of course, one could resist this conclusion – *if* one is willing to sever the close connection I have been taking for granted between "natural" and "explanatorily valuable". Perhaps that option could be profitably pursued; myself, I think it gives the game away. I will continue to take for granted that whatever else nature's joints are, they had better turn out to be distinctions that it is of the first explanatory importance to know about. And so, given the failure of Lewis's account of how large-scale joints are grounded in fundamental ones, it makes sense to turn to the theory of explanation itself for clues to an alternative.

3. Causal explanation and explanatory depth

Lewis's insight, summarized in the thesis Causal Explanation, is surely correct: to explain why an event occurs must involve giving information about its causal history. The reasons for this verdict are, I think, fairly obvious. For when we reflect on the abstract structure of our judgments concerning what causes what, and of our judgments concerning what explains what, we find that they are remarkably similar – too similar for this to possibly be a coincidence. In particular, in both domains we draw a firm distinction between events knowledge of which serves as a *good predictive basis* for other events, and events that *cause* or *explain* those other events; and we draw these distinctions in *exactly the same way*. For example, if Billy throws a rock at a window, and we

[8] Saturated hydrocarbons are molecules composed of hydrogen and carbon, where all chemical bonds are single covalent bonds (either carbon-carbon or carbon-hydrogen).

know that the rock is sufficiently hefty, very well aimed, and thrown with sufficient force, and that nothing stands poised to intercept it en route to the window, then we have an excellent basis for predicting that the window will break. But whether the throw explains the subsequent breaking, or is a cause of the subsequent breaking, depends on what is going on in the surrounding environment – and depends on those goings on in exactly same way. Thus, if Suzy *also* throws a rock at the window, and her rock gets there first, then Billy's throw neither explains nor causes this breaking, notwithstanding the fact that knowledge only of it would have allowed us to predict that the window would break.

These observations are, I take it, perfectly familiar, and it is largely because of them that causal theories of explanation are so dominant in the contemporary literature, having long since supplanted the logical empiricist deductive-nomological model, and having successfully resisted the incursions of various other accounts that downplay the importance of causation to explanation.[9]

All the same, there are legitimate and serious grounds for dissatisfaction, at least with Lewis's version of a causal theory of explanation. It will be instructive to highlight three, and then sketch some ways in which Lewis's account might be augmented and polished so as to deal with them.

First, the account remains far too schematic, without an account of what causation itself is. I say this, not out of some absurd notion that a philosophical theory that makes use of concept X thereby incurs an obligation to include an analysis of X; rather, there are reasons specific to causation that make this concept a poor choice of primitive, in an account of explanation. One is that causal relations between events at one scale are, very plausibly, *metaphysically grounded* in causal relations at smaller scales – and, ultimately, metaphysically grounded in the bare facts about the world's total history of complete physical states, together with the fundamental laws that dictate the evolution of those states. We would therefore deepen our understanding of what explanation is if we understood how this grounding works. More importantly, an account of explanation ought to make it clear why the acquisition of explanatory information is *valuable* for creatures like us – the sort of enterprise it makes

[9] I have in mind, for example, Kitcher's unificationist account (1989), and accounts that lean on probabilistic dependence – e.g., (Salmon, 1971).

sense to invest quite a lot of effort into. (This is a point that Jim Woodward has very effectively emphasized in his terrific recent work on explanation (Woodward, 2005).) And that sort of clarity will only be achieved via an account of causation itself. Finally, a review of the rich philosophical literature on causation will reveal that there are too many unanswered questions about causation for us to feel comfortable that we have an adequate grip on what the causal history of an event *is*. Is causation invariably transitive? Is causation by omission the same thing as ordinary causation? What about causation by double-prevention? Are the most basic causal relata really *events*, or should they rather be taken to be *facts*? All of these questions, and more, remain up in the air. (See (Hall & Paul, forthcoming) for extensive discussion.)

The second reason for dissatisfaction is that there is more – much more – to explanation than merely the explanation of particular events. In fact, in mature scientific inquiry, it is only very rarely that the explanation of particular events takes center stage. (E.g., a cosmologist might have as her life's work explaining the Big Bang; but you won't find many more examples like that.) Now, Lewis is perfectly aware of this fact, and makes no pretensions to having provided a complete philosophical account of explanation. Still, it is overwhelmingly plausible that the project of explaining particular events bears interesting connections to the other sorts of explanatory projects that scientists *do* put at center stage; and we shouldn't be satisfied with Lewis's account until it is developed in such a way as to make these connections clear.[10]

The third reason for dissatisfaction is that Lewis's account misses something of great importance to us when we seek the explanation of a particular event. An example will illustrate. A window has broken. Why? Because Suzy threw a rock at it. Now, we could obviously fill out that answer in many ways, thereby increasing the amount of explanatory information conveyed: we could trace the intermediate causes connecting Suzy's throw to the breaking; we could trace her throw's own causal origins; we could highlight the other causes contemporaneous with her throw with which it conspired in order to bring about the breaking. All of these ways of adding explanatory content Lewis's account, of course, recognizes. But it misses a distinct dimension along which our explanation of the window's breaking can be deepened. For what we might do instead is to highlight those aspects of Suzy's

[10] (Lewis, 1986b) includes some sketchy remarks on this topic. They do not suffice.

throw *in virtue of which it was able to bring about the breaking*, distinguishing them from other aspects that were causally (and so explanatorily) *irrelevant*. For example, the mass of her rock was important, but its color, not so much. And we could go further still: we could articulate, even with some mathematical precision, the structure of the way in which the window's breaking depended upon such factors as the rock's mass, the angle and velocity of the throw, and the distance between Suzy and the window. Granted, in any ordinary context all of this would be overkill – but that is not the point. The point here is that we have a kind of information that is clearly explanatory of the window's breaking, and that Lewis's account misses.[11] Again following Strevens (2009), I will say that this sort of information adds to an explanation's *depth*.

Let me now sketch an attractive approach – and one that has, I think, proved enormously philosophically fruitful over the last couple of decades – to understanding causation and explanation. It will go a fair way to answering the foregoing complaints. The central idea is extremely familiar: we should understand causation in terms of counterfactual dependence.

Now, this idea needs to be developed in the right sort of way, and two points in particular are critical to keep in mind. First, one should not think that the proper route to a counterfactual theory of causation is by way of some all-purpose semantic account of ordinary language counterfactuals (as Lewis himself apparently did; see his 1979). No, the counterfactuals in question need to be specialized. My own view is that they should have the following archetypal

[11] There is a complication, because Lewis in various places (for example 1986d) advocated a theory of events according to which at least some of the information being discussed here could be imported into the individuation conditions for the events themselves. Thus, he distinguishes events that are perfectly coincident in space and time on the basis of which of their features are essential, and which accidental. So perhaps you could say something like this: There were many throws that took place, of all of which Suzy was the agent; one of these was, inter alia, essentially a throw of a rock with such-and-such mass, but only accidentally a throw of a rock with such-and-such color. This throw caused the breaking. Other of the coincident throws – e.g., the throw that was essentially of such-and-such color a rock, but only accidentally a throw of a rock with such-and-such mass – did not. So you might hope – at least, if it's really important to you to preserve the exact letter of Lewis's account – that the sort of information that I am suggesting contributes to the explanatory depth of an event-explanation can simply be coded into the exact specification of the events that make up the target explanandum's causal history. I rather doubt this can be done, and I'm certain it cannot be done without producing a philosophical theory of events that is unpleasantly cumbersome. But at any rate, it doesn't really matter. The crucial point for the purposes of our discussion is simply that a good account of event explanation needs to recognize, somehow or other, the dimension of explanatory goodness that I've highlighted.

form: if the state of the world at time t had been just as it actually is, except with respect to goings on in a particular localized region, and if the state local to that region had differed in such-and-such a way, then the state of the world at a certain *other* place and time would have differed in such-and-such a way. There is a clear story to tell about how counterfactuals of this form are underwritten by the fundamental laws of nature; see for example (Maudlin, 2007b) or (Hall, 2011b). The basic idea is quite simple: given the alteration to the time-t state of the world specified in the antecedent, one simply updates the entire counterfactual history by plugging this state into the fundamental laws. Thus, this recipe shows how the fundamental laws, together of course with the totality of facts about our world's history, endows our world with a rich *localized dependence structure*.

The second point – which we will mostly set aside for the remainder of this essay – is that it will not do to simply *identify* causal structure with localized dependence structure. That is the lesson of cases of preemption, as for example the case mentioned earlier, in which Billy and Suzy both throw rocks at a window with deadly accuracy, but Suzy's rock gets there first. Here we see near-perfect symmetry between the relations of localized dependence holding between Suzy's throw and the breaking, and between Billy's throw and the breaking; but for all that, there is a striking asymmetry in causation. The right response, in my view, is not to try to exploit the tiny discrepancies in localized dependence structure that distinguish Suzy's throw from Billy's, but rather to recognize that part of what we're tracking when we track causal structure is the intrinsic structure of the processes that connect causes to effects. But this story gets quite complex, and at any rate, when told correctly, it still vindicates the thought that in ordinary cases, causal structure is relatively cleanly manifested in localized dependence structure. (Compare: we know better than to identify the property of *being disposed to Φ under conditions C* with the counterfactual property of being such that you would Φ if you were in conditions C; cases of masking and mimicking refute that simple equation. All the same, in ordinary cases the dispositional property is indeed manifested in the simple counterfactual behavior.)

So suppose that – while, again, bracketing the real and important worries raised by cases of causal preemption – we identify the sort of causal structure that is relevant to our explanatory interests with localized dependence structure. Then the sources of dissatisfaction mentioned above go away. We have an account of causal/explanatory structure that shows how this structure

is grounded in features of our world that are metaphysically more fundamental. Without going into details, the account can be used to say sensible things about thorny topics such as causation by omission, putative counterexamples to transitivity of causation, etc. (See (Hall & Paul, forthcoming).) We can sketch simple and attractive reasons why causal/exploratory information, so construed, is valuable to creatures like us: it is the sort of information in light of which we are able more effectively to navigate our world – a point emphasized by Woodward (2005) – and in addition it is the sort of information that creatures with our limited epistemic capabilities can reasonably hope to acquire, and by means of which we can reasonably hope to build up a more and more sophisticated understanding of the nomological structure of our world (Hall, 2011a).

We can also draw a connection between our prosaic practice of explaining ordinary events and the more refined and exalted explanatory aims of the sciences: whereas, in developing an explanation for why some particular event occurred, we are aiming to spell out one fairly *restricted* bit of the world's overall localized dependence structure, one of the central aims of the sciences is to discover *wide reaching* and *nomologically robust generalizations* concerning this structure, and patterns within it.

Finally, an account of explanatory depth falls out rather naturally. Consider our case of the broken window. By saying that it broke because Suzy threw a rock at it, we are conveying a bit of information about the localized dependence structure within which the breaking of the window is embedded. But we are doing so only very crudely: we're saying, roughly, that if the region of space in which her throw took place had differed just enough so that she didn't throw (but, say, stood idly by), then, given the laws, the entire state of the world would have evolved forward in such a way that the window did not break. We convey much more sophisticated information about the given localized dependence structure if, instead, we detail *which variations on Suzy's throw would or would not have led to a breaking*. Talk of the features of her throw in virtue of which it was a cause of the window's breaking is really just talk of the contours of such variations.

4. A puzzle about explanatory depth

But at this point, a puzzle emerges. Exploring it will take us directly back to the issues discussed in §1. To bring out the puzzle, we will draw again on our example of the broken window.

Here, in the abstract, is a way to see what we're doing when we increase the "depth" of our explanation of the window's breaking, by not merely citing the fact that Suzy threw a rock at it, but by articulating which counterfactual variations on her throw would and would not have led to a breaking. We have picked out a certain region of space-time: the region in which Suzy's throw takes place. Holding fixed the state of the rest of the world at the given time, the state of that particular region is, in the actual situation, such as to lawfully guarantee that the window breaks (at a certain time).[12] There are a multitude of nomologically possible alternatives to the exact physical state that this region instantiates. Some of these alternatives are such as to *still* lawfully guarantee the window's breaking (at roughly the same time, and, again, holding fixed the state of the rest of the world at the initial time); some are not. In aiming for explanatory depth, it appears that we are aiming to show how exactly the distinction between the former sorts of alternatives (example: an alternative in which the color of the rock is different) and the latter sorts (example: an alternative in which the rock is substantially lighter) is to be drawn.

But that can't be right, for we are doing some something more, and something much more subtle. Consider that one of these nomologically possible alternative states of the given region is the following: Suzy has no rock in her hand, but is in the process of running up to the window to level a vicious kick at it. Clearly, when we try to deepen our understanding of why the window broke by asking which sorts of variations on Suzy's throw would still have led to a breaking, we do not mean to include this scenario as one of them. Why isn't this alternative relevant, in the specific sense that it should be classed together, for explanatory purposes, with such alternatives as the one in which Suzy throws a rock of a slightly different color?

You might think the answer obvious: the actual cause of the window's breaking is a *throwing of a rock*, whereas whatever is going on in the imagined alternative – call it a "preparing to execute a running kick" – is a *different sort*

[12] Note that we are assuming determinism here. The story is more complicated if we relax that assumption, but not in ways that it would be profitable to explore.

of event altogether, and so cannot be seen as a "variation" on the actual cause. But to say that is to do nothing more than to highlight that we have already somehow managed to impose a certain scheme for taxonomizing events as the *explanatorily appropriate* one. It is, evidently, acceptable for our explanatory purposes to classify events as "throws" or as "preparations to kick", but not appropriate to use "disjunctive" classifications such as "throw or preparation to kick". (For if it were appropriate, then the counterfactual scenario in which Suzy is preparing to execute a running kick *would* count as a variation on the actual scenario; after all, she's *doing the same sort of thing*; just in a different way.) Why not? It seems to me we have not *answered* the original question, so much as forced it to take a different form.

Can we simply draw on the natural/non-natural distinction (really: the *more* natural/*less* natural distinction), at this point? Perhaps as follows: A way of classifying events that lumps together Suzy's actual throw with (inter alia) her counterfactual preparations-to-kick draws a much *less natural* distinction than a way of classifying that simply lumps together her actual and counterfactual throws; and it is *for that reason* that we achieve explanatory depth by deploying the latter classification, but not the former. Or, to put the point in terms of similarity, a counterfactual preparation-to-kick is *too dissimilar* to the actual throw, as compared to counterfactual variations on this throw, to count as one of the alternatives among which we need to distinguish, in order to achieve explanatory depth. (And these similarity facts, in turn, are grounded in the facts about the less-than-perfectly-natural properties instantiated in the actual and counterfactual scenarios.)

But this sort of appeal to the more/less natural distinction strikes me as far too cavalier. Given the problems raised in §1 for explaining what this distinction comes to, appealing to it doesn't *illuminate* so much as *label* what we are trying to understand. And at any rate, the presupposition that we achieve explanatory depth by focusing on those nomologically possible alternative states of the given region of spacetime that count as *variations on the actual throw* – aiming to distinguish those of them that lead to a breaking from those that do not – is false. Consider a variation in which Suzy throws the rock with slightly bad aim, just missing the window – but throws the rock so hard that it breaks the sound barrier, with the subsequent sonic boom shattering the window. We do not mean to classify this variation, *either*, together with variations in which we merely ring changes on the rock's color, etc. And this, notwithstanding that it is an alternative that clearly counts *as a*

variation on her throw. Again, why not? Not because lumping this variation together with the others produces an overly "unnatural" classification.

There *is* an answer, of course: breaking a window by way of a sonic boom counts, as compared with breaking it via direct impact, as a sufficiently different *way* of causing it to break, that our explanatory purposes are ill-served by lumping them together. That answer is correct, as far as it goes. But, again, it's a dead end to think that you can unpack that answer in some philosophically illuminating fashion by claiming that what *makes* our explanatory purposes ill-served in this way is that the classification in question is *insufficiently natural*.

And now for a radical suggestion: it's not just that this move leads to a dead end, it's that it gets things *exactly backwards*. What makes a classification that blends breakings-via-impact together with breakings-via-sonic-boom unnatural is that it ill-serves our explanatory purposes. In the remainder of this essay I am going to explore this idea.

5. Unification as a cognitive aim

Let's recap. A very good idea about causal explanation is that what we are seeking, when we set out to acquire such explanations, is information about particularly distinctive features of, or patterns in, the structure of localized counterfactual dependence that our world exhibits. One sort of structure in particular is the kind of structure knowledge of which gives our causal explanations "depth": it is the structure constituted by facts about how the localized state of the world in one place and time counterfactually covaries with the state in another place and time. Put another way, we are not merely interested in knowing that what goes on *here* and *now* causes what goes on *there* and *then*; we are also interested in tracking how this causal relation remains stable under counterfactual variations in the cause. But we are, it has emerged, not interested in just *any* old variations: for it appears to be bad explanatory policy to track variations in the cause under which its causal relation to the effect still obtains, but in an *overly different manner*. And so we have arrived at the need to understand what these distinctions among "ways of causing" are themselves grounded in.

What I have suggested is that it is a mistake to turn to the more/less natural distinction for help, because that distinction itself needs to be grounded in an account of what makes for better or worse satisfaction of our explanatory purposes. I will argue for that suggestion indirectly, by outlining a way that our

explanatory purposes could be served that could plausibly serve as grounds for a more/less natural distinction, and by leaving it a pregnant open question how *else* this distinction could be grounded.

It would be going in a circle, at this point, to give an account of our explanatory aims that merely returned to the themes discussed in the last section, asserting that these aims consist in the acquisition of causal information about the target explanandum. That is, of course, *one* of our aims, and a crucially important one at that; more generally, it seems to me that there is no way that we as scientific inquirers can come to an adequate understanding of our world, without knowledge of the metaphysical dependency relations (causal or otherwise) that knit it together. It's just that that cannot be the *whole* philosophical story about explanation. And what we need, at this point in the dialectic, is precisely the other part of the story. I therefore propose that one of our aims, in trying to develop an understanding of our world, is, in addition, to *develop cognitively effective means for organizing our information about the world*, in particular causal information.

That idea has clear connections to what, in the philosophical literature on explanation, has gone by the name of "unificationism" – which is, principally, Kitcher's unificationist account of explanation (1989). On this account, very roughly, explanations are arguments that instantiate very widely applicable patterns of argument. For reasons best left offstage, I do not think Kitcher's account succeeds, so let me hereby alert you that I do not in any way mean to be drawing upon it. I appropriate the label "unificationist" simply because, like Kitcher, I think that one important part of what we are after in explanation can be accurately (if very incompletely) described as the acquisition of a *unifying picture of the world*.

Now, the idea that one of the central things we are after in explanation is the development of cognitively effective means for organizing our information stands desperately in need of development itself. I do not have a theory to offer of just what a "cognitively effective means of organizing" is (and not, alas, merely for reasons of lack of space). But it is easy enough to find evocative examples that, I think, do an extremely effective job of bringing out the unificationist strand in our thinking about explanation. Here is one that is slightly goofy, but for all that one of my favorites.

Consider the following initial segment of an infinite sequence of natural numbers:

1,1,1,2,3,2,1,3,5,4,2,5,7,8,3,7,9,16,5,11,11,32,8,13,13,64,13,17,...

Perhaps you've figured out the rule that generates the sequence. Perhaps, on the other hand, you find it confusing. You don't *understand* it. You don't know *why it has the form it does*. If so, the following way of reorganizing the initial segment will make things crystal clear:

1,	1,	1,	2,
3,	2,	1,	3,
5,	4,	2,	5,
7,	8,	3,	7,
9,	16,	5,	11,
11,	32,	8,	13,
13,	64,	13,	17, ...

Looking down the columns, we see that the sequence is just an interleaving of the odd numbers, powers of 2, fibonacci numbers, and prime numbers. Once you see this, you understand the sequence. But not by acquiring a special sort of information about it. (The sequence is, after all, not the sort of thing that has "causes", or that "metaphysically depends" on anything else.) To me, examples like this evoke in its purest form the idea that to understand some subject matter is to organize one's information about it in the right sort of way.

Not surprisingly, examples with this particularly clear character – in which explanatory insight is achieved not at all via the provision of a special sort of dependency information, but entirely by organizing the information we have in the right sort of way – are much easier to find in mathematics than in the sciences, simply because in mathematics the only kind of dependency information that's available is information about logical entailment, and that only gets you so far, explanatorily speaking. Just consider the fact that mathematicians routinely distinguish proofs that are illuminating from proofs that aren't; and yet the unilluminating proofs are, for all that, proofs! So something else must ground the distinction. I suggest that the something else concerns how the illuminating proofs generalize to other results, how they highlight easily overlooked connections between their subject matter and other mathematical topics, and so on; in short, they are illuminating to the extent that they contribute to the effective organization of mathematical knowledge.

In the sciences, by contrast, explanation almost always involves the provision of interesting, distinctive dependency information, and for that reason it can be difficult to see that unificationist requirements on understanding also play an important role. Still, some examples bring out these requirements rather nicely. Consider the periodic table of the elements, which

is unquestionably of immense value to us in enhancing our understanding of the chemical and atomic behavior of atoms, and which has this value *precisely* because of the brilliantly effective way in which it organizes our information about this behavior.

So let's grant that explanation even in the empirical sciences involves both distinctively metaphysical and distinctively psychological aspects: on the one hand, we want a special sort of information – information about what depends on what, metaphysically speaking – but on the other hand, we want our information, especially our dependency information, to be organized in the right sort of way – where what makes for good organization, presumably, depends on potentially quite idiosyncratic features of human psychology.[13] Then the next thing to notice is that what makes for good organization is very often going to be a holistic matter.

Consider, again, the periodic table of the elements. To be sure, what makes this such a powerfully effective tool for understanding is in part that is a table of the *elements*: and the distinction between elements and non-elements does not obviously involve any holistic considerations. (Rather, it seems that we focus, for explanatory purposes, on elements simply because they are highly stable configurations of matter, and so the sorts of things about which it is possible to make useful generalizations concerning their behavior.) But it is the *periodic* table of the elements because of the way in which it classifies elements into different chemical types. And what makes the particular scheme of classification built into the table so explanatorily superior to the multitude of logically possible rivals cannot, I think, be appreciated by examining its components piecemeal. It is not, as it were, that a certain amount of explanatory goodness attaches to any scheme that distinguishes noble gases from things that are not noble gases, and a certain additional amount of goodness attaches to any scheme that incorporates a distinction between metals and non-metals, and so on; with the overall goodness of our own scheme simply being the sum of these individual goodnesses. No, it is because of the way in which our scheme *as a whole* arranges our knowledge of the chemical and atomic features of the elements that it is so explanatorily powerful.

[13] Well, maybe not. Maybe, indeed, the very possibility of rational thought requires that understanding be achieved partly by the imposition of a priori principles of organization. If you're obsessively concerned to preserve the pure, unadulterated objectivity of our explanatorily valuable classifications, that might be the way to go.

I think that something very similar happens – albeit in a fashion that is less rigorous, and much more difficult to discern – in our thinking about causation and causal processes. We saw above that we distinguish a breaking-via-sonic-boom as a sufficiently different way of breaking a window from a breaking-via-impact that it would be a bad mistake, given our explanatory aims, to count the latter sort of breaking as a variation on the former. It is enormously philosophically tempting to think that this distinction must be grounded entirely in features specific to the two kinds of breaking: just by, as it were, closely inspecting paradigm examples of the two kinds of breaking, one would be able to see that our explanatory aims require us to distinguish them *as* separate kinds.

But I think that is a mistake. I think it is much more plausible that in coming to grips with the vast profusion of causal processes we encounter even in ordinary life, we very early on (and almost certainly unconsciously) hit upon certain schemes for organizing these processes into types. Now, to borrow an idea from Lewis's work on laws of nature – see (Lewis, 1983b) and also (Loewer, 1996) – two extremely important desiderata we impose on candidate schemes (again, not consciously!) are, plausibly, the following:[14] First, it's good for a candidate scheme to be *simple*, not necessarily in the sense that it includes a small number of types, but perhaps in the sense that it makes use of a small number of basic parameters to characterize those types. Second, it's good for a candidate scheme to have the resources needed to express *powerful, informative* generalizations about causal structure (in particular, the sorts of generalizations that populate the special sciences, and that in the philosophical literature typically go, misleadingly, by the name "ceteris paribus laws"). These desiderata work in tension: consider that one way to get a simple scheme is to let the sole type of event be "event" and the sole type of causal process to be "causal process". Whatever you said about the causal structure of the world, by means of this scheme, could be said quite simply. But not very informatively. It is immensely plausible that achieving the best balance between these desiderata will involve holistic considerations.

Now, maybe it's hubris to think that the schemes we humans have developed are the best possible, for purposes of effective organization.

[14] Not the only two, surely. For example, it's plausibly a desideratum that a scheme for taxonomizing events and causal processes not yield up kinds whose membership is difficult or impossible for creatures like us to empirically determine.

(Certainly, one of the striking effects that a good scientific education can have on one is opening one's eyes to the availability of very different schemes, that sometimes improve dramatically on our ordinary ones.) But I will suppose that they are good enough. At any rate, what I would like to suggest is that lumping the two kinds of breakings together is an explanatorily bad move *not* solely because of features intrinsic to paradigm instances of each, but because no scheme for organizing the vast amount of information we possess about the causal structure of our world that did so could possibly meet the desiderata on effective organization as well as the scheme we have arrived at.

6. Conclusion: mapping the large-scale joints of the world

Let's return now to the questions left over from section 1. How is it that the natural/non-natural distinction, as it appears at scales above the most mereologically fundamental, is determined by the fundamental physical structure of the world (the structure given by the pattern of instantiation of perfectly natural properties, together with the fundamental laws of nature)? How does the structure of reality at the most fundamental levels determine the map of reality's joints at less fundamental scales? The right answer, I think, is that it doesn't – at least, not alone. The picture is rather the following: given how the world is fundamentally (where I take this to include: how its fundamental laws are), the world has a perfectly definite localized dependence structure, which for purposes of keeping things simple (i.e., ignoring the complexities that cases of preemption introduce) we will take to just *be* its causal structure. But this structure does not, as it were, come equipped with a uniquely best way to describe it, even at a given scale. Rather, we impose on it various taxonomies – different ones for different scales, certainly, and sometimes even different ones at the same scale, given that it can sometimes be useful for us to highlight certain patterns in the world's localized dependence structure at the expense of others. These taxonomies sort events and the causal processes that knit them together into kinds, and do so subject to the constraint that the sorting provide us with maximally effective tools for organizing our view of the causal structure of the world at the given scale. The map of the large-scale joints of the world is just constituted by whatever distinctions figure in such optimal taxonomic schemes.

It follows that, in a certain sense, the distinction between more and less natural properties at larger mereological scales fails to be perfectly objective:

for this distinction is determined in part by which taxonomic schemes do the best job for creatures like us of providing tools for the efficient and effective representation of causal structure. So it would be a bad mistake to think that what we are doing when we investigate the large-scale structure of the world is merely discovering the natural distinctions that are there to be drawn. But it would be just as bad a mistake to think that how to draw these distinctions is somehow entirely up to us. To say that is simply to forget that the fundamental physical structure of the world – which, I'm supposing, is what it is quite independently of facts about the structure of human cognition – is also an indispensable ingredient. In sum: It is a complex interplay between purely objective facts about reality's physical structure, on the one hand, and psychological facts about the structure of human cognition, on the other, that grounds the "joints" that nature exhibits at large scales.

Now, just in case this point wasn't obvious, what I am offering is *not* a proper theory of the more/less natural distinction. What *Lewis* offered, in his proposal that the naturalness of a property is fixed by the length of the shortest canonical predicate expressing it, was a proper theory. (Granted: it slips from "proper theory" back to "approach" if we amend it by saying that simplicity of the predicate also matters, while leaving it vague how simplicity itself is to be measured, and how simplicity and length trade off.) What I have offered are remarks that point in the direction of a theory. My hope is that they point, at least, in the *right* direction. At any rate, they pretty clearly cry out for elaboration.

Three avenues in particular seem to me worth pursuing. First, the picture I've sketched needs input from empirical psychology, since that is where we can hope for insight into how it is that organizing schemes in fact function in human cognition. Second, it would be helpful to explore how our taxonomizing strategies work when applied to toy models – Conway's game of "life", say.[15] Third (and relatedly), it would be helpful to explore case studies from especially well-developed and mature special sciences – organic chemistry, say.

It's highly unlikely that the results of such inquiry will yield anything as pristine as Lewis's account. No, it's going to be messy – and, maybe, messy in case-specific ways. For example, the way the natural/nonnatural distinction

[15] See for example http://en.wikipedia.org/wiki/Conway's_Game_of_Life.

plays out in organic chemistry may not be the same as the way it plays out in, say, evolutionary biology. But that's to be expected, if indeed this distinction results from an interplay between facts about physical structure and facts about human cognition in the way I have suggested.

REFERENCES

Hall, N. (2010). David Lewis's Metaphysics, *Stanford On-Line Encyclopedia of Philosophy.*

Hall, N. (2011a). Physical and metaphysical possibility. Manuscript.

Hall, N. (2011b). Causation and the sciences. In S. French & J. Saatsi (Eds.), *Continuum Companion to the Philosophy of Science*. London: Continuum, 96–119.

Hall, N. & Paul, L.A. (forthcoming). *Causation and the counterexamples: A user's guide.* Oxford University Press.

Kitcher, P. (1989). Explanatory Unification and the Causal Structure of the World. In P. Kitcher & W. Salmon (Eds.), *Scientific Explanation.* Minneapolis: University of Minnesota Press, 410–505.

Lewis, D. (1979). Counterfactual Dependence and Time's Arrow. *Noûs*, *13*, 455–476. Reprinted with Postscripts in (Lewis, 1986a), 32–66.

Lewis, D. (1983a). *Philosophical Papers, Volume I.* Oxford: Oxford University Press.

Lewis, D. (1983b). New Work for a Theory of Universals. *Australasian Journal of Philosophy*, *61*, 343–377. Reprinted in (Lewis 1999), 8–55. Page references are to this latter printing.

Lewis, D. (1984). Putnam's paradox. *Australasian Journal of Philosophy*, *62*, 221–236.

Lewis, D. (1986a). *Philosophical Papers, Volume II.* Oxford: Oxford University Press.

Lewis, D. (1986b). Causal Explanation. In (Lewis, 1986a), 214–240.

Lewis, D. (1986c). Events. In (Lewis, 1986a), 241–269.

Lewis, D. (1994). Humean Supervenience Debugged. *Mind*, *103*, 473–490.

Lewis, D. (1999). *Papers in Metaphysics and Epistemology*. Cambridge: Cambridge University Press.

Lewis, D. & Lewis, S. (1970). Holes. *Australasian Journal of Philosophy*, *48*, 206–212. Reprinted in (Lewis 1983a), 3–9.

Loewer, B. (1996). Humean Supervenience. *Philosophical Topics*, *24*, 101–127.

Maudlin, T. (2007a). *The Metaphysics Within Physics*. Oxford: Oxford University Press.

Maudlin, T. (2007b). A Modest Proposal Concerning Laws, Counterfactuals, and Explanation. In (Maudlin 2007a).

Salmon, W. (1971). Statistical Explanation. In W. Salmon (Ed.), *Statistical Explanation and Satistical Relevance*. Pittsburgh: University of Pittsburgh Press: 29–87.

Sider, T. (2011). *Writing the Book of the World*. Oxford: Oxford University Press.

Strevens, M. (2009). *Depth: An Account of Scientific Explanation*, Cambridge, MA: Harvard University Press.

Woodward, J. (2005). *Making Things Happen: A Theory of Causal Explanation*, Oxford: Oxford University Press.

Any Sum of Parts which are Water is Water

Henry Laycock †
laycockh@queensu.ca

ABSTRACT

Mereological entities often seem to violate 'ordinary' ideas of what a concrete object can be like, behaving more like sets than like Aristotelian substances. However, the mereological notions of 'part', 'composition', and 'sum' or 'fusion' appear to find concrete realisation in the actual semantics of mass nouns. Quine notes that 'any sum of parts which are water is water'; and the wine from a single barrel can be distributed around the globe without affecting its identity. Is there here, as some have claimed, a 'natural' or 'innocent' form of mereology? The claim rests on the assumption that what a mass noun such as 'wine' denotes – the wine from a single barrel, for example – is indeed a unit of a special type, the sum or fusion of its many 'parts'. The assumption is, however, open to question on semantic grounds.

1. Innocence, Guilt, and the Utterance of Quine

1.0 *Mereology.* Mereologists posit a variety of contentious principles of composition, whereby diverse objects – wholes, 'fusions' or 'sums', analogous to sets but without a membership relation – may be constructed on the basis of specified ranges of objects, abstract or concrete, assigned the roles of 'parts', parthood in this context being akin to set-theoretical inclusion. The question of whether, in any particular axiomatized system, the definitions can be somehow plausibly mapped into any natural-language understandings of 'object', 'whole' and 'part' is a further question, as is the question of whether there ('really') are objects, recognisable independently of the mereological system, which actually satisfy its axioms.

† Department of Philosophy, Queen's University, Canada.

Naturally, the mereologist is free to deny that her favoured system is contentious; she may urge that fusions of the kind contrived within her system do exist, that the system does in fact match up with the real world, or part of it. She may even go so far as to declare that her system posits no novel, hitherto unsuspected entities – that its principles of composition and fusions are in effect implicit in our everyday world-view, or in the natural-language use of words like 'and' and 'part', the use of plural referring expressions, and so on.

David Lewis, for example, declares his own system of mereological constructions to be ontologically innocent in some such sense (Lewis, 1991). Nevertheless, there is a serious question as to whether his system is indeed as he declares it to be. As Byeong-Uk Yi has plausibly argued, it is *not* innocent to propose that as Lewis has defined them, there are such things as sums or fusions of individual objects, and it is *not* innocent to treat composition as identity (Yi, 1990).

I hereby declare my sympathy for Yi's robust sense of the constitution of reality, and take the view that along with round squares, doctrines such as that of the Trinity are simply incoherent.[1] Nevertheless, matters look very different, once attention is re-directed from such Lewisian objects as the sum of Tom and Jerry to the mereological status of such kinds of stuff as water, wine and bread. For as it happens, there is here a prominent and *prima facie* plausible, genuinely innocent application of mereological principles, observed among others by Quine in *Word and Object* (Quine, 1960, p. 93).

1.1 '*Quantities*'. Quine there remarks that the natural-language semantics of what he calls *mass terms* directly satisfy mereological principles: 'any sum of parts which are water is water' is the way he puts it. And Helen Cartwright, in her influential doctrine of 'quantities', has followed Quine in this regard, writing that «there is, I think, a 'natural' mereology for a given set of quantities of, e.g., water in the sense of 'quantity of –', as I have elsewhere tried to explain» (Cartwright, 1975a).

[1] I here mean the '3-in-1' doctrine as initially formulated and voted in by the Homoousian majority among the 1800 bishops at the First Council of Nicaea in C.E. 325, whereby three individuals are pronounced a single substance. As for Yi, and as Russell observes in another context, «a robust sense of reality is very necessary in framing a correct analysis of propositions about unicorns, golden mountains, round squares, and other such pseudo-objects» (Russell, 1919, pp. 169–170).

Now suitably defined, concrete mass terms denote material *stuff* or *matter* – stuff like water, iron, hydrogen and wine.[2] And since much of our world consists precisely of such material stuff or matter, there appears to be a solid basis for the direct, non-constructed relevance of mereological principles to a substantial part of our actual world. No independently contrived mereological system seems required – mereological principles apply directly to mass term semantics. Indeed, there here seems to be no more 'innocent' alternative to what looks like an inherently mereological system.

In a recent article, Keith Hossack also, rather cautiously, endorses just such a view (Hossack, 2000). Hossack adopts a terminology not of 'quantities' but of 'masses' (a term coined for anything which might be referred to as *some stuff* – some wine, some bread, some water, and so forth – and proceeds to cautiously remark that 'mereology is perhaps most successful in the case of masses'.[3] He continues

> If we read 'x is part of y' as 'x is some of y', then transitivity holds. The lower half of the water is part of the water in the glass, which is part of all the world's water, and the lower half of the water is part of the world's water. Moreover the axioms of fusions or sums seem to hold at least if the mass is pure; any arbitrary collection of masses of water does seem to have a unique sum.

In effect, then, what Hossack sees in the Quinean formula is precisely a real world vindication of fundamental mereological principles.

[2] The definition in (Cartwright, 1975b) is narrow and precise.

[3] The term 'masses' is due not to Cartwright but to Dean Zimmerman (1995). The intent behind the use of such technical terms is to 'convert' a non-count noun to a count noun (CN, for short) – in effect, to assert (with or without argument) that what the non-count term denotes is, contrary to its natural language grammar, a single object, individual or thing. It is then crucial for the purpose at hand to clearly distinguish the wine in a glass or bottle – that is, what is interpreted to be the maximal mass of wine in the glass or bottle, in this technical sense, from the 'mass' of wine presented by a glass or bottle, in the more ordinary non-gimmicky and innocent sense of a single compact and continuous body of stuff, a body or mass which is itself divided into and replaced by three smaller masses, when the wine is poured. In this more ordinary sense, bottles and glasses of wine are themselves masses of wine – we drink bottles and glasses of wine – and these are compact, individuated bodies of the liquid, dependent for their individuated existence on the containers which constitute their 'forms'. In this sense of 'mass', the wine is distinct from the mass or masses it happens to be in; it was in a single mass and it, the very same wine, is now in three distinct masses. The employment of a natural-language term for what is in fact a distinct technical purpose in this manner is regrettable, in persuasively eliding or obscuring the very considerable difference in the implications of the two associated concepts.

Hossack's remark is cautiously hypothetical, and the fact that he writes 'x is *part* of y' rather than 'x is *a part* of y' is itself an indication of his caution.[4] Nonetheless, Hossack speaks of *masses*, in the plural, and treats references to water, like Quine, as references to distinct objects, each of which might constitute *a* part of other objects of the kind. And it is evident that such things behave more like sets and their subsets than like Aristotelian substances. But in the very nature of the case, and regardless of whatever else might be true, the ideas of parts and wholes are at the very least the ideas of discrete objects – units, individuals or things. Correlatively, theories of parts and wholes are *theories* of individual objects, individually countable parts and wholes. Whether or not each part itself has proper parts, as with Lewisian gunk, each part must at any rate be *one*.[5]

2. Innocence Lost

2.0 *Real scattered objects*. Let us consider these matters more closely. I choose to speak of wine instead of water, and begin with the bottle of Brunello on our dining table. The bottle at first contains a certain amount of wine, 75 ccs, we may suppose.[6] Having opened the bottle, I pour the wine into three glasses to prepare for lunch. The wine which previously occupied a single compact region of space, defined by the inner surface of a bottle, is identical with the wine which is now spatially distributed, in multiple glasses and in multiple locations around the table; *some of* the wine which was in the bottle is now in each of the three glasses. Plainly, both the degree of 'scatter' of the wine, and the number of glasses which are used to contain it, are entirely irrelevant to its identity.[7] What, if anything, might seem interesting or

[4] In speaking in this way, Hossack seems to acknowledge something special in the 'some of' relationship – something underpinning his use of the expression 'part' without the singular determiner.

[5] On gunk – a recent term for an ancient concept, and one which corresponds to the actual semantics of what I have elsewhere called 'pure' non-count nouns – see (Lewis, 1991).

[6] We may naturally and innocently refer to this wine as 'an amount of wine'; but the formal behaviour of terms with this structure is complex, to be treated with caution.

[7] The question of the relationship between the wine which now occupies three distinct glasses, and the wine in each one of the glasses, is addressed in section 3.

remarkable about this state of affairs? One answer to this question would proceed as follows.

Our day-to-day experience of and interaction with the world accustoms us to thinking primarily in terms of 'ordinary objects' or Aristotelian substances – structured, physically organised objects, the parts of which bear determinate spatial relationships to one another. The fact is, however, that there seems to be something here of a very different nature – something capable of occupying spatially disconnected locations, where the degree of separation and number of distinct locations have no bearing whatsoever on its identity. If this is correct, it is surely interesting, even perhaps remarkable, in itself. It is as if we have stepped outside the abstract axiomatic constructions of mereology, to encounter a real-life demonstration of something resembling mereological wholes and parts, before our very eyes. It is tempting to say that – so far as the non-atomic *semantics* of words like 'wine' are concerned – the wine which was once in the bottle *really is* no more than the totality of its potentially or actually scattered parts.[8]

Now there are two key points in all of this. There is what I take to be an indisputable *fact*, that the identity of an amount of stuff, unlike that of a concrete individual substance, is independent of its degree of scatter or spatial distribution. On the other hand, there is a mereological *interpretation* of the fact – in this particular case, an interpretation of the relationship which exists between the wine on the table, and *some of* it (that is, the relationship which exists between *some wine* and *some of* some wine). Or, what comes to the very same thing, there is a mereological interpretation of the formal status of what expressions like 'the wine on the table' and 'some wine' actually *denote*. The significance of the indisputable fact itself remains to be addressed; I first consider each of these equivalent interpretations in its turn.

2.1 *The relationship*. Consider now two neutral ('innocent') descriptions of two aspects of this state of affairs. (i) The wine from the bottle just is – is straightforwardly identical with – the wine in glasses A, B and C. (ii) The wine in *each* of the glasses A, B and C is *some of* the wine in glasses A, B and C; hence the wine in glasses A, B and C is an amount of wine which is – and here,

[8] The semantics of 'wine' require that whatever is some of some wine will also be some wine.

is or consists of – a lesser amount of wine in glass A, a lesser amount in glass B, and a lesser amount in glass C.[9] Furthermore, *the* amount of wine in glasses A, B and C is indisputably the (mathematical) sum of the amounts of wine in each of the glasses A, B and C.[10] The semantical phenomenon of cumulative reference corresponds, I suggest, to the relationship represented by *is or consists of* in statement (ii), rather than to the relationship of identity as expressed in statement (i).[11] Intuitively, the relationship represented by *is or consists of* must in some sense be one of composition, not the (straightforward, pure, 'innocent') relation of identity.

This, no doubt, is how the relationship is understood in the first instance by the mereologist – as a compositional part/whole relationship between one sum object and three part objects. She may also, with Lewis and the Trinitarians, guiltily take it to be a relationship of identity; and in my view, there is a sense in which she might even be right about this. But if so, she is right, only because there are neither at least three proper part-objects on the table, nor one maximal whole-object.[12] *Either* it is a relationship of identity, and there are no wine-objects on the table, *or* it is instead a relation of composition, and there are at least four such objects on the table (and most likely, countless such things).[13] That is the view which I wish to now explain and defend.

[9] I do not say that the wine in the glasses is or consists of three amounts of wine; the grammar of 'amount' licenses no such assertion; see note 15.

[10] 'An amount of wine' is an equivalent concrete natural-language designation for 'some wine'; 'the amount of wine' is an abstract natural-language designation for the universal measure of some wine; the wine in different glasses might yet be the same amount of wine (there might be 25ccs of wine in each glass). The grammar of the expressions 'an amount of ___' and 'the amount of ___' closely parallels that of the grammar of the expressions 'a number of ___' and 'the number of ___'; the former is used to make concrete indefinite reference to a number of objects, the latter is an abstract reference to a number. There are major differences between numbers and amounts – the question of a choice of measures does not arise for numbers, giving them a certain 'absolute' status.

[11] In referring to the wine in glasses A and B and C, I have not referred to the wine in any one of the three glasses; and in referring to the wine in each of the three glasses, I have not referred to the wine in all three glasses.

[12] The thought that deity appears both as one and as many, but in reality is neither, might conceivably have a certain pantheistic appeal.

[13] For the first disjunct, the relationship of identity – non-standard though it is – obtains between a (non-singular) amount of wine and itself (see note 15). The second disjunct, which Lewis would no doubt accept, is simply false, or so I urge in 2.2 and 2.3.

Now the mereologist is likely to share the common assumption that the object–concept itself is an all-purpose, all-inclusive concept – that whatever we may speak of, refer to or think about *cannot fail* to be an object in some minimal but reasonably precise sense. As Russell writes in a well-known passage, «whatever may be an object of thought, or may occur in any true or false proposition, or can be counted as one, I call a term. This, then, is the widest word in the philosophical vocabulary» (Russell, 1903/1937, p. 43).

But this view can itself be questioned. If it is an essential feature of the object-concept that objects must be capable of being numbered and distinguished, then the fact that there exists a major semantic category of *non-count* nouns (*NCNs*) – nouns precisely for the diverse varieties of *stuff* – should constitute grounds for re-evaluation of the mereologist's fundamental but unexamined belief.[14] There are several arguments for the conclusion that the basic subject-matter of the mereological thesis in this domain – that is, whatever is *some wine* – is *not*, as the mereologist supposes, a unitary object.

2.2 *The denotation (i).* In 'On denoting', Russell declares that '*the*, when it is strictly used, involves uniqueness'. This is Russell's criterion for a *singular* description – where such a description is understood to be a description which *denotes*, if anything, a single unit of some kind. The essence of the Theory of Descriptions is given by this declaration, which says in effect that if 'the' is joined with a *singular* noun 'F', then 'the F' means 'the *only* F'. That is, a description having the form 'the F' is singular, if the concept-expression it contains (the 'F' itself) applies or purports to apply *uniquely*. Here, I attempt only to explain the rationale behind this theory; I have defended it at some length elsewhere.[15]

Now if there is exactly one fish on a certain table, then the concept-expression 'fish on that table' can be said to apply uniquely, and the definite description '*the* fish on that table' counts as semantically singular.[16] Here, the noun 'fish' *itself* has a semantically singular occurrence. Suppose however that there are many fish on the table – some fish on this plate, some others on that. Then 'fish on the table' applies to the fish on this plate, and *also* to the fish on that plate. Here, the use of '*the* fish on the table' to denote is evidently *not* a

[14] The matter is addressed in (Laycock, 2010).
[15] The criterion is defended at length in Laycock, 2006.
[16] I choose the noun 'fish' because it has the syntax of a zero-plural noun.

singular use. It is, of course, plural (the *fish-es*), and the noun 'fish' itself has a semantically plural occurrence. The fish on this plate are now *some fish* which are *some of* the fish on the table.

Consider then the use of the description 'the wine on the table', in the presence of two glasses of wine. Here, the concept-expression 'wine on the table' applies to *both* the wine in this glass and the wine in that glass. It follows that – much as with 'the fish on the table' in its plural use – '*the* wine on the table' cannot be semantically singular. However, unlike 'fish' in its count noun sense, 'wine' has no *other* semantic form – it lacks a cognate singular form, and so cannot be plural *either*. Although non-plural, the NCN is akin to a plural CN in being semantically non-singular; and *qua* semantically non-singular, it cannot designate a (single) object.

2.3 *The denotation (ii).* It follows immediately from an understanding of the count/non-count contrast that what underlies the kinship of plural CNs and NCNs is indeed the fact that both are semantically *non-singular*.[17] CNs, first, are semantically *either* singular or plural: singularity and plurality are the twin semantic sub-categories which jointly exhaust this category of nouns. It would seem then to follow directly that the category of NCNs can be *neither* singular nor plural (a fact which itself is obscured if, instead of *non-count noun*, the appellation *mass noun* is employed). NCNs are then to be classed as semantically non-singular, simply in virtue of being non-count. And given this, two propositions follow automatically.[18]

First, NCNs have in common with plural nouns the distinction of being semantically non-singular. The semantic kinship between NCNs and plural nouns is these days widely recognised; what is typically unnoticed, in this recognition, is the simple fact of its non-singular semantic *basis*. And second, NCNs have in common with singular nouns the distinction of being semantically non-plural. Quantification involving such nouns must then also be semantically non-singular, a fact reflected in their non-acceptance of singular determiners. As with plural nouns, we speak of 'all water', 'some water' and

[17] To the best of my knowledge, this claim was first advanced in (Laycock, 1975). The kinship of NCNs and plural CNs is noted, among others, by Schein (1994). Schein however argues that any formalisation must involve «reduction to singular predication, via a Davidsonian logical form». But since NCNs are neither singular nor plural, this is impossible.

[18] These semantic points are argued in (Laycock, 2006), see in particular chapters 1, 3 and 4.

'more water', but not in the singular of 'a water', 'each water' or 'one water'; and 'any', 'all' and 'some' appear to interact with NCNs much as they do with essentially plural nouns.

Now as it happens, this non-singular analysis of the semantics of NCNs has also been affirmed by Tom McKay.[19] In a helpfully concise account, McKay notes that while NCNs are indeed on a par with plural nouns in respect of their non-singularity,

> Plural discourse has natural semantic units that are the same as those of singular discourse, but stuff discourse has no natural semantic units, and reference and predication seem to proceed on a different model than that of an individual and a property. (McKay, 2008, pp. 310–311).

In consequence, he urges that in the case of words like 'water',

> We should not expect a successful reduction to singular reference and singular predication, something that the application of traditional first-order logic would require [...] when we say that water surrounds our island [...] our discourse is not singular discourse (about an individual) and is not plural discourse (about some individuals); we have no single individual or any identified individuals that we refer to when we use 'water'.

There are, in a word, *no individuals* introduced by the use of 'water', and to this extent, McKay and I are in complete agreement.

3. The Ontological Insignificance of an *Amount* of Wine

3.0 *The relationship of being 'some of ___' that which is 'some ___' once again.* It is a truism well worth repeating, that the ideas of parts and wholes are the ideas of discrete objects-units, individuals or things. Individual units and their unitary parts are uncontroversially ubiquitous. Every fish is such a unit, and its eyes are parts of it. Here, we have three units – a fish, and each of its eyes. Now suppose there are exactly ten fish on the table, on three plates. That fish which is closest to me is *one* of them. But it is not *a part* of them, because while *it* is a unit, *they* are *not* a unit. They are *ten* units, and nothing can be a

[19] See his 'Critical Notice' of Laycock 2006 (McKay, 2008). McKay is the one and only philosopher to date to have written a book devoted entirely to non-singular predication and reference, though his subject matter is almost entirely that form of non-singular predication which is plural (McKay, 2006).

part of something else, unless both it and what it is a part of are individual units. Similarly, it is not a *member* of them, and so does not 'belong' to them; it may be a member of one or another club, gathering or other ensemble, but *they* themselves are not, as such, identical with a club or gathering or any other unit of that genre. They are many, while the gathering to which they earlier belonged was one.

Furthermore, any two of the fish cannot be a part of the fish on the table, because any two of them are no more a unit than are all ten of them. Any two of them are *some of* them, and the (innocent) relationship which holds between *a number of objects* and *some of* those objects – between *the fish* on the table and *some of* those fish, for example – is a different kind of relationship from that of (innocent) whole and part, or of set and subset.[20] The distinctive nature of the 'some of' relationship is recognised by Lewis among others, although it is plainly not the same relationship (innocently speaking) as that of part to whole. To think of two of them as being (somehow 'literally') *a part* of them, is of course to think of both 'two of them' and 'them' as naming *units*. And indeed if this is what we think, then at least for the purposes of the current issue, we do indeed have objects of a very different kind – a different *category*, in fact – from 'ordinary' substantial physical objects – objects which have (potentially or actually) scattered parts. At the same time, however, we have lost our unperverted contact with reality, and must be deemed ontologically guilty in Yi's good sense.

Much as the fish on this plate are *some of* the fish on the table, so the wine in this glass is *some of* the wine on the table. But in *neither* case do we have something which can be innocently counted either as *a unit* or *a part* of something else.

3.1 *The ontological status of quantity.* Both the idea of an amount of matter, and the idea of a number of objects, combine the ontic categories at issue – the categories of *objects* and of *matter* – with a notion of *quantity*, a notion of how many or how much. The idea of *a number of* objects – if it can be thus dignified

[20] We may decree that the fish are members of a set having cardinality of ten. But insofar as a set is itself a unit whole, *they* cannot be identified with this whole, it being one and they many – unless, of course, we are Descatarians who believe that ten (ordinary) individuals might be not only ten different and distinct (ordinary) individuals but as well be (identical with) only and exactly one (unusual) individual, ten times as big as any one of the ten.

– self-evidently combines the neutral idea of 'objects' *simpliciter* with the further idea of determinate but unspecified *multiplicity* or number. Similarly, the idea of an amount of stuff combines the neutral idea of stuff or matter *simpliciter* with the further idea of determinate but unspecified amount. And non-singular *references* to either matter or to objects necessarily incorporate this fact. Thus, the bare plural sentence

There were fish on the table during lunch

says *less* than the non-bare sentence

There were a number of fish on the table during lunch.

The latter carries implications of *identity* – it might be followed by

No-one ate any of *them*

– but the former carries no such implications; it might be followed by the pseudo-anaphoric

They were constantly replenished by the waiters.[21]

And in parallel identity-related fashion, the bare non-count sentence

There was wine on the table during lunch

says less than the non-bare sentence

There was an amount of wine on the table during lunch.

The former sentence might continue 'Prosecco to begin with, and Brunello to follow up' (a continuation which would be bizarre indeed for the latter sentence).

Syntactically, the form of plural reference involves a single *grammatical* subject, whose *semantic* character involves a determinate number of distinct units – units whose identities are drawn together, merely via the collective *form* of a single human act of reference. The idea of *a number of* objects involves, in effect, the *fusion* (ordinary sense) of the *ontic* category of objects

[21] By a 'pseudo-anaphoric' relationship, I mean that unlike standard anaphora or cross-reference, in which a pronoun picks up the identity of a previous reference, no such identity, no sameness of fish, is implied in the use of the bare noun.

with the *semantic* category or form of plural reference.[22] Hence the answer to the question of just what *category* the phrase 'a number of objects' itself expresses or represents must be that it represents, if anything, an essentially *hybrid* category – one which reflects the semantical category of plural reference itself. It follows that a sentence, the subject of which has this semantic form, lacks any distinct metaphysical significance on precisely this account. This non-ontic character of plural reference is reflected, among other things, in the fact that a number of objects cannot be said to have *ceased* to be until the last one of them has ceased to be; while those objects cannot be said to have *persisted*, or retained their identities, unless all of them have persisted.

There are in short no such *things* as 'numbers of objects'; there are individual objects, and there are numbers. From the standpoint of assertions of existence, the sole categorially or ontically salient fact consists in the information that there are *objects* of this or that kind which are thus-and-so, in a given context. In expressing empirical quantity, and thereby laying the semantic basis for plural identity-statements, the presence of 'some' constitutes the introduction of an element which is *adventitious*, from the standpoint both of the relevant kind and of the ontic category itself. The non-singular 'some of' relationship, along with the terms which it relates, is a hybrid relationship without ontological content. What it is *not* is an ontic relationship of part to whole.

Essentially these points apply also to the idea of an amount of stuff or matter, and to the relationship between that and the neutral idea of matter *simpliciter*. There are no such *things* as amounts of stuff; there is stuff of one sort of another; and there are amounts. Indications of quantity are a matter of empirical information, information which is has no bearing on the categorial or ontic import of a sentence. In postulating entities where none exist, Quine, Cartwright and Hossack stand together in the dock, to be pronounced ontologically guilty.[23]

[22] As such, the category of objects is without a number – it is 'neutral', neither singular nor plural, but it may be represented or expressed in either singular or plural form.

[23] Occam's presence among the jury is unnecessary.

4. Innocence Regained

Now I began by noting that when there is wine in three glasses, *there is something* which occupies spatially disconnected locations, where both the degree of separation and the number of distinct locations are completely irrelevant. But while this is undoubtedly the case, in just what sense *is* this the case? In what sense is there something in spatially disconnected locations? Once the non-singularity of NCNs is clearly understood, the sense in which this is the case is neither remarkable nor interesting – or rather, no *more* interesting or remarkable than the actual semantics of NCNs themselves.[24] At the end of the day, it is no more interesting or remarkable than the fact that, if there are fish scattered about in various locations, then there will be something which occupies – that is, there will be some *things* which *occupy* – these spatially disconnected locations.

In this latter case, although *objects* are *scattered*, there are no scattered *objects*. An ontologically innocent or neutral account of this state of affairs has the scatter distributing merely over *many*, rather than being a collective feature of some *one* – there being no such 'one'. The scatter is a feature of plurality; there is no unitary plural whole with many individual parts; there are simply many individuals, along with *references* to all of those individuals collectively, or to some of them in particular. This then is the 'innocent' or ordinary view of fish and of references to fish. There are merely objects of this kind, distributed in different places.

Likewise, so far as the wine on the table is concerned, while there is an amount of stuff which occupies these spatially disconnected locations, that stuff is no more a unit than are the fish; so that although the *stuff* is scattered, here too there is no scattered *unit*. Rather, there is merely stuff of this kind, distributed in different places and in varying amounts; here, scatter distributes not over the many but the much. It is, first and foremost, the direct illusion of unity which generates the belief that there are mereological entities before our very eyes, entities which then appear to legitimate the mereological posit in a way it would otherwise be lacking.

[24] For a devotee of the semantics of the predicate calculus, the semantics of NCNs should seem remarkable indeed.

And finally, for the cases of both stuff and things, there are genuine wholes which are the innocent sums of innocent parts. Much as the amount of wine in glasses A, B and C is an abstract *amount* – not an amount of wine – the (mathematical) sum of the amounts of wine in each of the glasses A, B and C, so the number of fish on the plates is a number, the mathematical sum of the number of fish on each plate.[25] Numbers and amounts, unlike wine and fish themselves, display an authentic mereological relationship of addition to one another.

By contrast, the fact that this fish and that fish are two is nothing other than the relationship of non-identity between them; numerical *adjectives* in general express no more than the non-identities of countable individuals.[26] It is perhaps tempting to think that there must *be* something in our concrete reality, to which ('abstract') numbers directly correspond. But number itself requires no physical 'embodiment' to have application to reality; arguably, all that is required is the mere existence of one-one correlations between objects – correlations which can be established, along with 'same number', without being able to count.

REFERENCES

Cartwright, H. (1975a). Some Remarks about Mass Nouns and Plurality. *Synthese, 31*(3–4), 395–410.

Cartwright, H. (1975b). Amounts and Measures of Amounts. *Noûs, 9*(2), 143–164.

Hossack, K. (2000). Plurals and Complexes. *British Journal for the Philosophy of Science*, *51*(3), 411–443.

Laycock, H. (1975). Theories of Matter. *Synthese, 31*(3–4), 411–442.

Laycock, H. (2006). *Words without Objects.* Oxford: Clarendon Press.

[25] What I am calling an abstract amount is a universal – if the amount of wine in two separate glasses is identical, then we have an amount of wine in one glass which embodies the same universal or abstract amount as the wine in the other glass.

[26] This is I suggest the innocent view of multiplicity, a view I have tried to defend in (Laycock, 2006).

Laycock, H. (2010). Object. In E.N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2010 Edition). <http://plato.stanford.edu/archives/fall2010/entries/object/>.

Lewis, D. (1991). *Parts of Classes.* Oxford: Blackwell.

McKay, T. (2006). *Plural Predication.* Oxford: Oxford University Press.

McKay, T. (2008). Critical Notice of *Words without Objects*. *Canadian Journal of Philosophy, 38*(2), 301–323.

Quine, W.V.O. (1960). *Word and Object.* Cambridge, MA: The MIT Press.

Russell, B. (1903/1937). *The Principles of Mathematics*. Cambridge, UK: Cambridge University Press.

Russell, B. (1919). *Introduction to Mathematical Philosophy*. London: Allen & Unwin.

Schein, B. (1993). *Plurals and Events*. Cambridge, MA: The MIT Press.

Yi, B.-U. (1999). Is Mereology Ontologically Innocent? *Philosophical Studies, 93*(2), 141–160.

Four Theses on the Alleged Innocence of Mereology*

Massimiliano Carrara†
massimiliano.carrara@unipd.it
Enrico Martino†
enrico.martino@unipd.it

ABSTRACT

In *Parts of Classes* David Lewis attempts to draw a sharp contrast between mereology and set theory and he tries to assimilate mereology to logic. For him, like logic but unlike set theory, mereology is "ontologically innocent". In mereology, given certain objects, no further ontological commitment is required for the existence of their sum. On the contrary, by accepting set theory, given certain objects, a further commitment is required for the existence of the set of them. The latter – unlike the sum of the given objects – seems to be an abstract entity whose existence is not directly entailed by the existence of the objects themselves. The argument for the innocence of mereology is grounded on the thesis of *composition as identity*. In our paper we argue that: (T1) arguments for the ontological innocence of mereology are not conclusive. (T2) Some arguments against the ontological innocence of mereology show a certain ambiguity in the innocence thesis itself. (T3) The innocence thesis seems to depend on a general conception of the nature of objects and on how the notion of ontological commitment is understood. Specifically, we think that the thesis is the manifesto of a realistic conception of parts and sums. (T4) Quine's notorious criticism of the set-theoretical interpretation of second order logic seems to be reproducible against Lewis'defence of mereology. To the purpose we construct a mereological model of a substantive

* The paper is based on Carrara & Martino, 2007. Thanks to Roberto Ciuni and Giorgio Lando for the helpful comments and suggestions to an earlier version of the paper.
† Department of Philosophy, University of Padua, Italy.

fragment of set theory, adequate to ground the set-theoretical semantics of second order logic.

0.

In *Parts of Classes* (Lewis, 1991) David Lewis attempts to draw a sharp contrast between mereology and set theory and to assimilate mereology to logic. He argues that, like logic but unlike set theory, mereology is "ontologically innocent". Consider the following sentences:

(1) There is a cat, Mina, which is sleeping.
(2) There is a mouse, Gino, which is dancing.

Whoever asserts (1) is committed to the existence of a cat whose name is Mina. Whoever, after the assertion (1), asserts (2) is committed to the existence of a mouse whose name is Gino. Whoever, after the assertion (1) and (2), accepts set theory is further committed to the existence of an entity – a set – whose elements are Gino and Mina. By contrast, if one accepts logic no further commitment is required apart from a commitment to Mina and Gino.

Lewis argues that the same is for mereology: given certain objects, no further ontological commitment is required for the existence of their sum (or fusion). By contrast, by accepting set theory, given certain objects, a further commitment is required for the existence of the set of them. The latter – unlike the sum of the given objects – seems to be an abstract entity whose existence is not directly entailed by the existence of the objects themselves.

The goal of this paper is to analyse arguments pro and cons the ontological innocence of mereology. We argue that:

(T1) Arguments for the ontological innocence of mereology are not conclusive.
(T2) Some arguments against the ontological innocence of mereology show a certain ambiguity in the innocence thesis itself.
(T3) The innocence thesis, apart from Lewis's defence, seems to depend on a general conception of the nature of objects and on how the notion of ontological commitment is understood. Specifically, we think that the thesis is the manifesto of a *realistic* conception of parts and sums.
(T4) The alleged innocence of mereology is subject to Quine's notorious criticisms of the set-theoretical interpretation of second order logic. To the purpose, we construct a mereological model of a substantive fragment of set theory, i.e. the one that grounds the principal model semantics of second order logic.

The paper is divided into six sections. In the first section we recapitulate Lewis's version of mereology. In the second one we analyze Lewis' argument for the innocence of mereology: an argument grounded on the thesis of *composition as identity*. Lewis analyses two different versions of the thesis: the first one is the *strong composition thesis* (StrongCom), according to which certain objects *are* their sum, where the use of "are" would mean that composition is *literally* identity. The second version is the *weak composition thesis* (WeakCom), according to which composition is *similar* or *analogous*, under some aspects, to identity. In the third section we analyse some arguments pro and cons (StrongCom) – specifically the arguments by Lewis, Yi, and Van Inwagen against (StrongCom) – and we argue for (T1) and (T2). In the fourth section we analyse arguments pro and cons (WeakCom). Specifically, we analyse arguments pro (WeakCom) given by Lewis. Again, some arguments pro (T1) and (T2) are given in this section. In section five we construct a mereological model of a substantive fragment of set theory, i.e. the one that grounds the principal model semantics of second order logic, and we argue for (T4). In the last section we give an argument for (T3).

1.

"Mereology", literally the "science or theory of parts", stands for theories analyzing the relation "... is a part of ...". There are different formulations of mereology depending on the language adopted. Due to the fact that we would like to consider Lewis's defence of the ontological innocence of mereology we propose his formulation of mereology, suited to point out some relevant aspects of the problem we are analysing.

Lewis treats mereology in a plural language, a language extending that of first logic, including singular and plural reference, singular and plural quantification (for an introduction to a plural language see Boolos, 1984). In such a language we consider logical terms:

(a) Plural terms, for example the pronoun "them", or plural variables, for example "X" as symbolic counterpart.
(b) Plural quantifiers, for example, "there are some things... such that".
(c) A special two-place predicate "... is one of ...". This predicate admits a singular term in its first place and a plural one in its second place.

By adding to this vocabulary the non-logical predicate, "... is a part of ..." we obtain a language rich enough to formulate mereology. By means of the

predicate "... is a part of ..." one could define sums (or fusions) and the overlapping relation.

(Def.1) y and x overlap if and only if there is a z such that it is part of x and part of y.
(Def.2) y is a sum of the X if and only if each of the X is a part of y and each part of y overlaps one of the X.
(Def.3) The X compose y if and only if y is the sum of the X.
(Def.4) x and y are disjoint if and only if they do not overlap.

Mereology consists of the logical consequences of the following axioms:

(Reflexivity) x is part (non proper part) of itself.
(Transitivity) If x is part of some part of y, then x is part of y.
(Unrestricted Composition) If there are some X there is a sum of the X.
(Uniqueness of Composition) If y and z are sums of the same X then y = z.

For example, the following theorem is a logical consequences of the above axioms:

(Theorem 1) If there are two objects, neither of which is part of the other, then there is something else that is not identical with either of them.

2.

Consider again the sentences:

(1) There is a certain specific cat, Mina, which is sleeping.
(2) There is a certain specific mouse, Gino, which is dancing.

Whoever asserts (1) is committed to the existence of a certain specific cat whose name is Mina. Whoever, after the assertion (1), asserts (2) is committed to the existence of a certain specific mouse whose name is Gino. Suppose that someone – after the assertion of (1) and (2) – asserts:

(3) There is a sum of the mouse Gino and the cat Mina, Gina.

Is she committed to the further existence of the sum of the mouse Gino and the cat Mina?

If we follow the Quinian motto (Quine, 1939, p. 708) that to exist is to be the value of the bound variables, the answer should be positive: since sums are values of bound variables, mereology is committed to the existence of the sum of whatever plurality of objects X, no matter how they are given and however they are heterogeneous.

Question: is the ontological commitment to the existence of the sum of the X a further commitment? Specifically, is the commitment to the existence of the sum of the cat and the mouse a further commitment besides the existence of the cat and the mouse? Lewis's answer is: no. One could answer that (3) is a logical consequence of (1), (2) and mereology, more precisely, of (Theorem 1). But such an answer is trivial. In fact it does not say anything at all about the *ontological* commitments of mereology.

Lewis's point is that with (3) we have not introduced a new entity:

> Given a prior commitment to cats, say, a commitment to cat-sums is not a *further* commitment. The sum is nothing over and above the cats that compose it. It just *is* them. They just *are* it. Take them together or take them separately, the cats are the same portion of Reality either way. Commit yourself to their existence all together or one at a time, it is the same commitment either way [...]. I say that composition [...] is like identity. The 'are' of composition is, so to speak, the plural form of the 'is' of identity. Call this the Thesis of *Composition as Identity*. It is in virtue of this thesis that mereology is ontologically innocent: it commits us only to things that are identical, so to speak, to what we were committed to before. (Lewis, 1991, pp. 81–82)

Lewis's argument for the innocence of mereology is the following:

> (P1) Composition – a many-one relation – is like identity.
> (P2) The commitment to sums is already presupposed in the acceptance of the objects that are summed.
> (P3) Nothing could be considered more ontologically innocent than the request to accept something identical to things already accepted.
> (P4) No other entities beyond sums of individuals are introduced in mereology
> *therefore*
> (C) Mereology is ontologically innocent.

For Lewis, the sum of certain objects is the very same objects: their sum is *them* and nothing more. Speaking of sums of heterogeneous and/or scattered objects might seem to be inappropriate. But mereology is not concerned with that: the generality of the theory does not permit to exclude certain sums for reasons concerning the nature or the location of the objects which are taken into account.

Lewis's argument rests on the thesis (P1) of *composition as identity*. What does it mean that composition is like identity? The answer depends on the reading of (P1) one accepts. In fact, there are two of them: a strong reading (StrongCom) and a weak one (WeakCom). For (StrongCom):

(StrongCom) The predicate "are" used for the composition relation is literally the plural for of the "is" of identity.

Formally:

$\forall X \forall y ((y \text{ is the sum of the } X) \rightarrow y = X)$

Those who accept (StrongCom) argue that the sum of some things is literally identical to those things: things are their sum, the sum is that things. If so, it is obvious that there is no further commitment to anything else apart from the commitment to parts. In this perspective the predicate "are" of composition is just a different form of the "is" of identity in the same way in which predicates "am" and "are" in sentences as:

(4) I am Pino.
(5) You are Dino.

are alternative forms of "is" in a sentence as:

(6) She is Dina.

According to the above thesis the cat Mina and the mouse Gino together are literally identical to their sum, Gina, even if none of them is identical to it.

In the second reading of (P1) – the weak reading of composition – the composition predicate is only analogous to identity. (WeakCom) is formulated in the following way:

(WeakCom) The predicate "are" used for the composition relation is *analogous* to the plural form of the "is" of identity.

The strength of Lewis's argument for the innocence of mereology strongly depends on the truth of (P1), i.e. on the truth either of (StrongCom) or of (WeakCom). In the next two sections we analyze some arguments pro and cons the two readings of the *composition as identity* thesis. Specifically, in the next section we analyze Lewis, Yi, and Van Inwagen's arguments against (StrongCom).

3.

Lewis formulates two arguments against (StrongCom) in Lewis, 1991 (p. 87). The first one concerns the difficulties for a generalization of the definition of identity. Given the definition of identity between singular individuals in the following way:

(IS) x = y =df. ∀Z (x is one of the Z ↔ y is one of the Z)

one could try to generalize it to the case of a plurality and a single individual obtaining:

(ISP) X = y =df. ∀Z (each of the X is one of the Z ↔ y is one of the Z)

But if y is the sum of the X (where X are two or more disjointed objects) there are some things – the X themselves – such that each of the X is one of them but y is none of them. *Viceversa*, there are some things – y itself – such that y is one of them but none of the X is. For example: let the X be Mina the cat and Gino the mouse and y their sum, Gina. Taken X for Z then each of Mina and Gino is one of the Z but Gina is not. On the other side, taken Gina for Z then Gina is one of the Z but neither Gino nor Mina is one of Z.

Yi (2001) has proposed an argument against (StrongCom) similar to Lewis's one. Consider, again, the cat Mina, the mouse Gino and their sum Gina. Given (StrongCom) and mereology one could say that:

(7) Gino and Mina are (identical to) Gina.

but:

(8) Gina is not identical to Gino.

and:

(9) Gina is not identical to Mina.

From (8) and (9) one obtains that:

(10) Gina is not identical neither to Gino nor to Mina.

Moreover, the predicate "...is one of..." could be extended to a predicate with singular places so defined:

t is one of u ↔ ∀X (t is one of (u and X))

So, we can say:

(11) Gina is one of Gina.

And, given (11) and (7),

(12) Gina is one of Mina and Gino.

But, by (10):

(13) Gina cannot be one of Gino and Mina.

Then, (StrongCom) is wrong.

Lewis's second objection concerns the indiscernibility of identical (InId) i.e.:

(InId) $\forall x \forall y (x = y \rightarrow \forall F (Fx \leftrightarrow Fy))$

where the third universal quantification is of second order and "F" is a predicative variable. Lewis argues:

> Even though the many and the one are the same portion of Reality, and the character of that portion is given once and for all whether we take it as many or take it as one, still we do not really have a generalized principle of indiscernability of identicals. [...] What is true of the many is not exactly what is true of the one. After all they are many while it is one. (Lewis, 1991, p. 87)

Consider the following example. Suppose that the number of the X is n, where $n > 1$. Then, the plural predicate "...are exactly n" should apply – given (InId) – to y too, but the number of y is one.

Wallace (manuscript) replies to both Lewis's arguments. On the first one she observes that:

$X = y$

in

(ISP) $X = y =_{df.} \forall Z$ (each of the X is one of the Z $\leftrightarrow$ y is one of the Z)

has a distributive reading, i.e. each of the X is identical to y, whereas when y is a sum of the X identity has a collective reading. With reference to the above example: Gina is not distributively identical to Mina and Gino, but it is collectively identical to them.

Problem: if we read identity collectively, it becomes a primitive notion, indefinable in terms of plural quantification, as Lewis has observed. Moreover, the crux is that collective identity is not a genuine many-one relation. For, to hold that the sum of the X is collectively identical to the X amounts to denying that the sum of the X is a genuine entity: speaking of the sum of the X would be nothing but a device for referring to the X collectively. On the contrary mereology, in particular in Lewis's use for the reconstruction of set theory as "megethology" (see Lewis, 1993), needs to consider sums as genuine objects.

One reply to Yi's argument is – again – the distinction between a *collective* and a *distributive* reading of conjunction. Consider a sentence as:

(14) Dino and Lino have lifted the piano.

and suppose that Dino and Lino have lifted the piano all together. For sure, in (14) the "and" is not used as a propositional connective. In fact if it is so, from (13) we could infer that:

(15) Dino has lifted the piano and Lino has lifted the piano.

Assume that the piano is too heavy to be lifted only by Dino or by Lino. One can conclude that (14) is true whereas (15) is false. If it is so then the "and" in (14) should function as a connective yielding a plural term "Gino and Pino".

Yi admits that the plural term "Dino and Lino", in the sentence:

(14) Dino and Lino have lifted the piano.

does not refer singularly neither to Dino nor to Lino. That does not mean it does not refer at all. Suppose that a plural term as "Dino and Lino" does not refer singularly to none of the two individuals, but that it refers plurally to both of them. Then, the mereologist could argue that the sum is identical to Dino and Pino.

If Yi thinks that the commitment to plural terms is ontologically innocent he should say, arguing in the same way, that mereology is ontologically innocent. When Yi argues that there is a sum whose name is Gina such that it is neither Dino nor Mina, the mereologist could reply that there is a plurality of objects, Gino and Mina, which is neither a cat nor a mouse. In other terms, even if the mereologist could accept Yi's conclusion that there are some things which are neither a cat nor a mouse, he could reply that they are a cat and a mouse.

An easy reply to the above objection is to say that who has lifted the piano is not the sum of Dino and Lino. It is an action that Dino and Lino take together, simultaneously, an action not involving the presence of a new entity.

Likewise, saying that Mina and Gino are a cat and a mouse is just saying one is a cat and the other is a mouse and it is not saying that a single entity is a cat and a mouse. Saying that the term "Mina and Gino" possesses, after all, a reference, even if it does not refer neither to Mina nor to Gino, does not mean that we are referring to a different entity from Mina and Gino; it simply means that the term – just because it is a plural term – does not singularly refer to one of them. It refers simultaneously to both of them. This plural reference does not commit us to the alleged entity Gina.

Wallace reply to Lewis's second argument, the indiscernability argument, is recovered, with substantial modifications, by Baxter, 1988. For Baxter a way

to maintain (InId) and (StrongCom) is arguing for two kinds of identity, a *strict* and a *loose* one. Baxter gives the following exemplification of the above distinction:

> Suppose a man owned some land which he divides into six parcels. Overcome with enthusiasm for [the denial of composition as identity] he might try to perpetrate the following scam. He sells off the six parcels while retaining ownership of the whole. That way he gets some cash while hanging on to his land. Suppose the six buyers of the parcels argue that they jointly own the whole and the original owner now owns nothing. Their argument seems right. But it suggests that the whole was not a seventh thing. (Baxter, 1988, p. 579)

A justification of (StrongCom) is to argue that to strictly count -- via a *strict* identity – the many is to loosely count – via a *loose* identity – the one.

> (BT) The whole is the many parts counted as one thing. (Baxter, 1988, p. 579)

Even if Baxter argues that (BT) does not deny the existence of the whole, but just the *additional* existence of the whole, it seems to us that this popular mood does not reify the whole. Baxter's example demonstrates a *weak* use of the sum, not involving the existence of it as an entity. It seems to be a use of sums similar to the one of sets in a sentence as:

> (16) The set of the Germans camping in Pinarella has cardinality six hundreds.

A sentence one can reformulate without the introduction of the notion of set, saying that:

> (17) The Germans camping in Pinarella are six hundreds.

Likewise, the sentence:

> (18) I have seen a flock of six geese.

can be rewritten in this way:

> (19) I have seen six geese.

so that (18) does not involve that "flock" stands for a certain specific entity.

For Baxter, speaking of the sum of the X would be just another way of speaking plurally of the X. Unfortunately, mereology does not have just this eliminative use of the sums, since *each* individual in mereology is the sum of its parts. Even referring to a single individual is referring to a sum. If there are individuals, there are sums too!

Van Inwagen replies to Baxter's in Van Inwagen, 1994. Consider Baxter's example of the land and its six parcels and express the fact that there are two parcels with a different size. We will use a quantified sentence with the following form:

(20) $\exists x \exists y \exists z (y < x \wedge z < x \wedge \neg G(y, z))$

where "<" stands for the relation "... is part of ..." and "G" stands for the relation "... has the same size of ...". On how many objects do we quantify? It seems that we must quantify on seven entities, because the first existential quantifier is exemplified by the whole. End of van Inwagen's argument.

One could reply by arguing that, using the plural language, (20) could be rewritten quantifying – singularly and plurally – just on the six parts:

(20*) Among the X two of them have a different size.

Formally:

(20**) $\exists x \exists y$ (x is one of the X $\wedge$ y is one of the X $\wedge \neg G(x, y)$)

However, it seems to us that it is possible to revive van Inwagen's criticism simply modifying his example. Suppose I would like to express the fact that the whole land is larger than each of its parcels. The singular variables range on every parcel of the land (included the very same land) so that:

(21) $\exists x \forall y (y < x \wedge y \neq x \rightarrow \neg G(x, y))$

A second kind of objection to (StrongCom) has been formulated by van Inwagen (1994). It is an objection concerning the very intelligibility of (StrongCom). Consider Lewis's sentences:

(22) It (the sum) is just them (the cats composing it).
(23) They (the cats composing it) just are it (the sum).

In a semi-formal way, using the plural language, one could translate (22) and (23) as follows:

(22') The sum y of the X is just the X.
(23') The X are just the sum y of the X.

For van Inwagen it is easy to observe that the "is" of identity is used in a correct way (from a syntactical point view) when there are singular terms on the right and left side of the relation. So for example, we say:

(24) Tully is Cicero.

(25) x is y.

Alternatively, in the natural language we use the plural form of identity: the "are" of (plural) identity. Such a term is used in a correct way (from a syntactical point view) when there are plural terms on the right and left side of the relation. So for example, we say:

(26) Fichte, Schelling and Hegel are German idealists.
(27) The X are Y.

Problem: what is the meaning of a sentence where the "is" and "are" of identity are placed by a singular term on one side and a plural term on the other? Of course, we can define both the singular and plural form of identity in terms of the relation "...is one of...". The singular one should be:

(IS) x is y =df. ∀Z (x is one of the Z ↔ y is one of the Z)

and the plural one:

(IP) X are Y =df. ∀z(z is one of the X ↔ z is one of the Y).

Problem: how should we define the "hybrid" form "is/are" in terms of "... is one of..." or in some other similar way such as the definition of identity in terms of overlapping? If we follow this train of thought, Lewis's tentative explanation with the sentence «the "are" of composition is, so to speak, the plural form of the "is" of identity» seems to be false: whatever one could mean by the "are" of composition, it cannot be the plural form of the "is" of identity because the plural form of the "is" of identity is the "are" of identity. Van Inwagen's conclusion is that (StrongCom) is unintelligible because the sentences exemplifying it are ungrammatical.

A first reply to van Inwagen's argument has been to argue that in natural language there are "hybrid" uses of is/are (the examples are in Wallace, manuscript). Consider, for example, the sentences:

(28) Two cups are a pint.
(29) One pint is two cups.
(30) One kilometer is thousand meters.
(31) Thousand meters are one kilometer.
...

Unfortunately, it is easy to reply that in these mixed uses the predicate in question is not really the identity one. One pint – differently from the cups – is a unit of measurement. A kilometer and a thousand meters are different

measurements (expressed by different numbers) of the same size. The above examples can be easily rephrased in the following way:

(28') Two cups have the capacity of one pint.
(29') One pint has the capacity of two cups.
(30') One kilometer and 1000 meters measure the same distance.
...

But the problem does not seem connected with the hybrid form. In a plural language a plural term could denote a singular individual. The formula:

(32) y is the X

naturally means that y is the only one X according to our definition (ISP):

(ISP) X = y =df. ∀Z (each of the X is one of the Z ↔ y is one of the Z).

The same result is obtained with an example taken from natural language. Suppose there is a bell with a German name written (Wuerms), and Pino says:

(33) There are some Germans.

The sentence is true even if just a German stays in the apartment. van Inwagen's criticism is better interpreted as a criticism to those who read (32) as:

(34) y is the sum of the X.

To say that y is the only X (when the X are n, with n > 1) does not mean that y is the sum of the X.

4.

The second reading of (P1) is (WeakCom) and says that composition is just *similar* or *analogous*, under some aspects, to identity. The composition relation is so formulated:

(WeakCom) The predicate "are" used for the composition relation is *analogous* to the plural form of the "is" of identity.

In this second reading of the thesis of *composition as identity* one is confined to argue for a certain similarity between composition and identity. Similarity has many aspects. The aspects of the similarity Lewis shows are the following:

(Unrestricted composition) Just as everything is identical to something, likewise

given anyway some X they compose something.

So, for example, there is no special condition Gino must satisfy for being identical to himself. Likewise, there is no special condition Gino and Mina must satisfy for composing something.

(Uniqueness of Composition). Just as there cannot be anything identical with two distinct objects, likewise there cannot be two distinct sums of the same objects.

For example, there cannot be two distinct things both identical to Gino. Likewise, there cannot be two distinct objects both composed by Gino and Mina.

(Ease of Describing Sums) Just as if you fully describe the thing x you fully describe something identical to x, likewise if you fully describe the X you fully describe their sum.

For example, you can fully describe the object identical to Gino describing Gino. Likewise you can fully describing the object Gina composed by Gino and Mina fully describing Gino and Mina.

(Spatial Coincidence) just as x and y have to occupy the same spatio-temporal region if the first object is identical to the second one, likewise y and X have to occupy the same spatio-temporal region if the first one is the sum of the second ones.

For example, if there is an object Gino in a certain place at a certain time, Gino exists in the same place-time. Likewise, Gina is in the same region occupied by Mina and occupied by Gino.

On the ground of the above analogies Lewis proposes a defense of the ontological innocence of mereology. In fact, from:

(1) There is a cat, Mina, which is sleeping.

and

(2) There is a mouse, Gino, which is dancing.

it follows that:

(1') There is something the cat Mina is.
(2') There is something the mouse Gino is.

For Lewis, from (1') and (2'), just considering "are" as a plural form of "is", it follows that:

(3') There is something the mouse Gino and the cat Mina are.

Where the "something" in (3') is – for Lewis – the sum, Gina. End of the argument.

Objections. First of all, the sentence (3') does not seem to be a consequence of (1') and (2'). If you mean "are" as the plural from of "is", from (1') and (2') follows:

(3'') There are the cat Mina and the mouse Gino.

But (3'') simply says that they both exist. In fact, we can paraphrase (3'') in the following way:

(3''') There is something the cat Mina is and there is something the mouse Gino is.

The problem is that (3''') does not entail (3'). Formally, from:

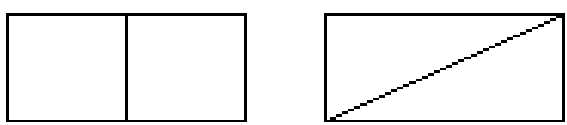
(35) $\exists x\,(x = a) \wedge \exists x\,(x = b)$

does not follow that:

(36) $\exists x\,(x = a \wedge x = b)$.

It seems to us that those who object that (36) is not the correct paraphrase of (3') – because in (3') one says that there is something Mina and Gino are collectively, not distributively – are wrong. In fact, the conjunction of (1) and (2) says that Mina and Gino are *each of them* something, and it does not say that they are collectively something.

Moreover, let us observe that there is an important disanalogy between composition and identity: while the description of an object identical to x describes x, the sum of the X does not describe the X at all. Given certain X and Y the sum of the X could be identical to the sum of the Y even if the X are not identical to the Y. Consider, for example, a rectangle A:



One could see the rectangle A as the mereological sum either of two squares (as in the left side of the figure above) or of two triangles (as in the right side). But if both squares and triangles were identical to their common sum, i.e. the

rectangle, then for the transitivity of identity, the squares should be identical to the triangles, that is absurd.

Besides, all aspects of the similarity at issue, apart from that concerning spatio-temporal regions, are applicable to the membership relation too. Nevertheless no mereologist would like to assimilate sums to sets.

The conclusive objection to Lewis's argument of the similarity is that it is just a *petitio principii*. Lewis's analogy between composition and identity rests just on the assumption of the existence of sums. For example, if one argues, in order to show an aspect of the similarity in question, that, as everything is identical to something, so too, however some X are chosen, they compose something, she presupposes just the existence of the sum of arbitrarily given objects.[1] But that is just what is in question and the similarity should demonstrate. Whoever challenges the ontological innocence of mereology denies the innocence of the alleged existence of sums of arbitrarily taken objects.

To conclude: arguments for the innocence of the mereology – both those based on (StrongCom) and those based on (WeakCom) – are not conclusive (our first T1 thesis). Moreover, we have argued that some arguments against the ontological innocence of mereology show a certain ambiguity in the innocence thesis itself. Some defences of the innocence seem to implicitly presuppose that the sum of certain objects X is not a genuine entity. Speaking of the sum of the X would be just another way of speaking plurally of the X. However, the relevant use of sums in mereology treats a sum as a single object (T2).

5.

Let us formulate a mereological model for sets of individuals. The goal of this section is to argue that the alleged innocence of mereology requires the ontological innocence of a substantive fragment of set theory, i.e. the one that grounds the principal model semantics of second order logic. Then, the ontological innocence of mereology is subject to Quine's notorious criticisms of the set-theoretical interpretation of second order logic.

[1] In Lewis, 1986 (pp. 211-213), Lewis gives a different argument for *Unrestricted Composition* based on vagueness. We do not consider it in this paper.

Let T be a theory of sets of individuals. The language L of T is a first order language with identity and with two kinds of variables:

x, y, z, ... variables for individuals;
α, β, γ, ... variables for sets;
$\in$ is the membership symbol.

Atomic formulas have the following forms:

$x = y$
$\alpha = \beta$
$x \in \alpha$.

Complex formulas are defined in the usual way. Axioms of T are:

Extensionality (ES) $\alpha = \beta \leftrightarrow \forall x\,(x \in \alpha \leftrightarrow x \in \beta)$
Comprehension (Com) $\exists\alpha\forall x\,(x \in \alpha \leftrightarrow A(x))$

where A(x) is any propositional function of L. It is possible to give the following mereological interpretation of T.

Let D be any domain of atoms (finite or infinite). Let D' be the sum of the atoms of D with a further atom j. Let us interpret the variables for individuals in the atoms of D and the variables for sets in the parts of D' containing j. Let us interpret $\in$ in the mereological relation < ("... to be a part of ...").

The presence of j has the effect of introducing the null set, j itself, and the singletons, the singleton {x} of x being the sum of x and j. And it is easy to verify the axioms (ES) and (Com).

(ES) Suppose that $\forall x\,(x \in \alpha \leftrightarrow x \in \beta)$. Then, α and β are sums of j and of the same atoms of D. So, they have the same atomic parts. So, they are identical.
(Com) Let A(x) be any propositional function. We must prove that there is a set whose elements are the individuals satisfying A(x). The searched set is just the sum of j and the atoms satisfying A(x) as (Unrestricted Composition) says.

So, assuming the existence of infinitely many atoms, we get a model of the power set of an infinite domain.

To the purpose of obtaining a mereological infinite model of T, the assumption of the existence of infinitely many atoms is replaced with that of the existence of infinitely many pairwise disjointed objects O (with or without atoms).

Let the objects O be interpreted as individuals. Let F(O) be the sum of the O and a further object j disjointed from each of them. Then the role of sets can

be played by the parts of F(O) containing j and not "cutting off" any of the O. That means that we take as set each part α of F(O) such that:

(i) α contains j

and

(ii) Each of the O either is a part of α or it is disjointed from α.

Observe that, using Lewis'definition of an infinite object, the existence of infinitely many disjoint objects follows from the existence of a single infinite object. For, consider the following definition:

Def (infinite). An object x is infinite if and only if x is the sum of some things, each of which is a proper part of another.

Given the above definition Def (inifinite) one can argue that:

(Theorem 3): If there is an infinite object there are infinitely many pairwise disjointed objects.
Proof. Let a be an infinite object that is the sum of some X each of which is a proper part of one of them. From X we can extract an infinite sequence of objects $b_0, b_1..., b_n,...$ such that each of them is a proper part of the subsequent. Then, objects b_1-b_0, b_2-b_1,..., b_{n+1}-b_n...(where b_{n+1}-b_n is the complement of b_n in b_{n+1}) are pairwise disjoint.

Let us observe that T is a substantive fragment of set theory, i.e. the one that grounds the principal model semantics of second order logic. Because of such ground Quine (1970) notoriously argues that second order logic is a wolf in sheep′s clothing. That means that second order logic is set theory in logic′s clothing.

Specifically, detractors of second order logic criticize the use of the comprehension principle (Com) as a *logical* principle. They hold that it does not possess the peculiar features of a logical principle. (Com) seems to concern the notion of set in the nowadays sense of set theory, where sets are understood as entities constituted by their elements. But, such a notion of set is highly problematic, and it does not seem to have a logic nature. Since it is possible to give a mereological interpretation of T, Lewis's assimilation of mereology to logic seems to be subject to the same objections (our T4 thesis).

6.

What about the ontological innocence of mereology? First of all Lewis' argument for the innocence of mereology shows a certain ambiguity in the use of the term "sum". On one side, Lewis seems to argue that, given certain objects X, referring to their sum is ontologically innocent because there is not a *new* entity as referent of the expression "the sum of the X". So, talking of the sum of the X is simply a different way of talking of the X, *looking at them as a whole*. This seems to be the only way to make intelligible, and plausible, the statement that:

(37) The X are their sum.

However, on the other side, Lewis's innocence is not understood as a mere linguistic use, where sums are not reified. It is not an innocence thesis comparable to that of plural reference where the reference to some objects does not require the existence of a single entity picking up them in a whole. Consider what Lewis says on this last issue:

> Plural quantification is innocent: we have many things, we do mention one thing that is the many taken together. Mereology is innocent in a different way: we have many things, we do mention one thing that is the many taken together, but this one thing is nothing different form the many. Set theory is not innocent. Its trouble has nothing to do with gathering many into one. Instead, its trouble is that when we have one thing, then somehow we have another wholly distinct thing, the singleton. And another, and another ,... *ad infinitum*. But that is the price for mathematical power. Pay it. (Lewis, 1991, p. 87)

In general, it is difficult to say what else could be the act of taking together the many as one if it is not an act of plural reference, an act that – according to Lewis – does not engage any singular entity apart from the many taken together. For sure, mereology and Lewis's use of it, specifically in his reconstruction of set theory as "megethology", requests that sums are taken as real objects.

Lewis seems to suggest that even if the sum of the X is a well determined individual, distinct from the X, the existence of such individual has to be necessarily accepted from whom has already accepted the existence of the X. In other words, committing oneself to the existence of the X would be an implicit commitment to some other entities and – among them – the sum of the X.

The problem is that arguing for this thesis implies a premise (P1) inadequate both to (StrongCom) and to (WeakCom): to the strong one (StrongCom) because the sum of the X is not literally identical to the X, to the weak one (WeakCom) because the analogy between composition and identity is – as we have argued – a *petitio principii*.

Moreover, we do not think that there are some conclusive arguments for the thesis that whoever accept the existence of the X is committed to the acceptance of the existence of the sum of the X (T1). For, since the parts of an infinite object constitute a non-denumerable infinity – for example the existence of natural numbers would imply the automatic existence of the continuum – such an argument would entail that there could be no infinity without a non-denumerable infinity. However, the thesis that any genuine infinity is a denumerable one has had some important advocates (see for example Kroeneker or Poincaré). So, a conclusive argument for the innocence of mereology seems to be highly implausible. This seems to be a general point about the alleged innocence thesis of mereology.

Last, we think that the thesis of the ontological innocence of mereology is the manifesto of a *realistic* conception of parts and sums. This conception consists of the following clauses:

> (i) Given any object x, it is well determined which parts it possesses; these are in turn objects whose existence is a necessary consequence of the existence of x.
> (ii) However any objects X are given, they automatically constitute a well determined object x which is their sum.
> (iii) We can refer singularly and plurally to parts and sums of given objects.

Obviously, one might wonder if such a conception is really ontologically innocent. One could object that it is not innocent because clauses (i)–(iii) are not. For example, clause (ii) could be considered as an ontological commitment to the existence of sums. But the innocence at issue does not concern the above-sketched conception. The innocence is embedded in the conception itself. In other words, someone who argues for clauses (i)–(iii) takes a point of view from which mereology *appears* to be innocent. Such a point of view forces us to consider the parts of any object as well-determined by the object itself and does not allow to separate the commitment to certain objects from that to their sum (T3).

REFERENCES

Baxter, D. (1988). Identity in the Loose and Popular Sense. *Mind, 97*, 575–82.

Boolos, G. (1984). To Be is to Be a Value of a Variable (or to Be Some Values of Some Variables). *Journal of Philosophy, 81*, 430–449.

Carrara, M., & Martino, E. (2007). *Sulla presunta innocenza della mereologia.* In A. Bottani & R. Davies (eds.), *Ontologie regionali.* Milano: Mimesis, 35–55.

Lewis, D. (1986). *On the Plurality of Worlds.* Oxford: Blackwell.

Lewis, D. (1991). *Parts of Classes.* Oxford: Blackwell.

Lewis, D. (1993). Mathematics is Megethology. *Philosophia Mathematica, 1*, 3–23.

Quine, W.V.O. (1939). Designation and Existence. *Journal of Philosophy, 36*, 701–709.

Quine, W.V.O. (1970). *Philosophy of Logic.* Englewood Cliffs, NJ: Prentice–Hall.

Van Inwagen, P. (1994). Composition as Identity. *Philosophical Perspectives, 8*, 207–220.

Yi, B. (1999). Is Mereology Ontologically Innocent?. *Philosophical Studies, 93*, pp. 141-160.

Wallace, M. (manuscript). On Composition as Identity. Webpage: http://www.unc.edu/~megw/OnCompAsId.doc.

Natural Properties, Supervenience, and Mereology*

Andrea Borghini†
aborghin@holycross.edu

Giorgio Lando‡
giorgio.lando@sns.it

ABSTRACT

The interpretation of Lewis's doctrine of natural properties is difficult and controversial, especially when it comes to the bearers of natural properties. According to the prevailing reading – the *minimalist view* – perfectly natural properties pertain to the micro-physical realm and are instantiated by entities without proper parts or point-like. This paper argues that there are reasons internal to a broadly Lewisian kind of metaphysics to think that the minimalist view is fundamentally flawed and that a *liberal* view, according to which natural properties are instantiated at several or even at all levels of reality, should be preferred. Our argument proceeds by reviewing those core principles of Lewis's metaphysics that are most likely to constrain the size of the bearers of natural properties: the principle of Humean supervenience, the principle of recombination in modal realism, the hypothesis of gunk, and the thesis of composition as identity.

* Earlier versions of this paper have been presented at the International Philosophical Symposium "Metaphysics, Language, and Morality" in Zagreb (December 1st–3rd, 2010) and at the conference in honor of David Lewis "Another World is Possible" in Urbino (June 16th–18th, 2011.) Some of the ideas originated during a reading seminar about "On the Plurality of Worlds" held at the Scuola Normale Superiore in Pisa in the academic year 2009–2010. We would like to thank all the people involved in these activities for their suggestions, and in particular Sergio Bernini, John Collins, John Divers, Mariano Giaquinta, and Massimo Mugnai. For comments on a more recent draft of the paper, we wish to thank Gabriele Galluzzo and Marco Nathan.

† College of the Holy Cross, Worcester, MA, USA.
‡ Scuola Normale Superiore, Pisa, Italy.

1. Are Natural Properties Instantiated by Minimal Entities?

Lewis's core doctrine of naturalness consists of three simple theses: first, some properties are natural; second, some properties are more or less natural than others; third, some properties – the perfectly natural ones – are more natural than all others. The interpretation of the doctrine, however, is far more difficult and controversial, especially when it comes to the bearers of natural properties. According to a certain reading of Lewis – the *minimalist view* – perfectly natural properties pertain to the micro-physical realm and are instantiated by "minimal entities": these are entities with a minimal size, that is without proper parts or point-like with regard to spatiotemporal extension, depending on the view. The present paper argues that there are reasons internal to a broadly Lewisian kind of metaphysics to think that the minimalist view is fundamentally flawed and that perfectly natural properties are instantiated at all levels, rather than only at the minimal one.

The minimalist view is not without *prima facie* textual support. The identification of the bearers of natural properties with minimal entities is indeed suggested by Lewis's preferred microphysical examples of natural properties. The most common are the charge and the spin of an electron,[1] where the latter seems to fit the role better than other subatomic particles just because it is expected not to be composed of smaller particles. Moreover, in some important passages Lewis characterizes so-called Humean supervenience – one of the most important principles of his entire philosophical work – in terms of "local" properties: «Humean Supervenience [...] is the doctrine that all there is to the world is a vast mosaic of local matters of particular fact, just one little thing and then another».[2]

The term "local", as it applies to properties, may be taken to mean that the natural properties involved in Humean supervenience are instantiated by entities located in space and time. But Lewis explains clearly that these properties are said to be "local" because they are *minimally* located in space and time; their bearers are points or point-sized entities: «We have geometry: a

[1] Electrons are considered as examples of perfectly natural properties in *On the Plurality of Worlds* (henceforth, OPW), p. 68, where Lewis is fixing his definition of perfectly natural properties. Just a few paragraphs above (p.64) Lewis also uses the example of unit positive charge, with some more reservation: «*let us assume* that unit positive charge is a perfectly natural property ...» (our emphasis).
[2] *Introduction* to the second volume of *Philosophical Papers* (henceforth, PPII), ix–xvi.

system of external relations [...] between points. And at those points we have local qualities: perfectly natural intrinsic properties which need nothing bigger than a point at which to be instantiated».[3]

In short, two sorts of evidences could lead to the conclusion that Lewis's natural properties are instantiated by minimal entities: the examples and the presentation of Humean supervenience. They are evidences of different sorts, resting on considerations that can be criticized by means of independent strategies. The examples stem from Lewis's naturalism and reductionism: it is up to natural sciences to identify natural properties; natural sciences can be reduced to physics; physics can be reduced to microphysics; the particles involved in microphysics (or in a core of microphysics to which microphysics can be reduced) are point-sized and have no proper parts. Still, a number of authors criticized the identification of the bearers of natural properties with minimal entities just from the point of view of contemporary physics, where the most basic entities which instantiate properties and enter relations are not always micro-particles devoid of structure, but are for example strings or fields. As a matter of fact some of these scholars (in particular Jonathan Schaffer, Andreas Hüttemann and Vassilios Karakostas[4]) see Lewis as a critical target. If Lewis's conception of science and of physics in particular was misguided, then Lewis's methodological principle that it is up to physics to identify natural properties could be retorted against his preferred examples and lead to the conclusion that also entities bigger than a point and endowed with proper parts are bearers of some natural properties.

The other sort of evidence does not seem to rest on better grounds. The problem of the size of the bearers of natural properties is connected with Humean supervenience. Yet the latter – to anticipate an argument offered in §4 – is compatible with the possibility that there are some properties that are at once natural, non-local, and excluded from the basis of recombination.

Looking at the situation from a different perspective, Lewis has several reasons to admit the existence of natural properties beyond Humean supervenience, the chief one being defining duplication and, hence, recombination. Indeed, given the prominent role covered by natural

[3] *Ibidem*

[4] See (Schaffer, 2003), (Hüttemann, 2004), (Karakostas, 2009). See also (Morganti, 2009) for a comprehensive survey of this debate.

properties, it is particularly relevant to look at Lewis's sources of evidence about the size of their bearers. Is there a formal, general criterion to identify those bearers? Is the criterion rooted in one or several theses characterizing Lewis's metaphysics, so that by adopting the theses we are *ipso facto* forced to conclude that the exclusive bearers of natural properties are minimal entities?

In this paper we shall review the core principles of Lewis's metaphysics that are most likely to constrain in some way the choice of the bearers of natural properties, in particular their size.[5] The purpose is analogous to that of Schaffer, Hüttemann and Karakostas, on one hand, but we are not going to draw on the results of contemporary physics. More in details, §2 introduces two alternative stances about the bearers of natural properties – called, respectively, minimalism and liberalism. §3 analyzes Lewis's concept of natural property, connecting it with the cognate notion of fundamental property in order to see if they place any constraint about the size of the bearers of natural properties. In §4 we come back in the same vein to Lewis's so-called principle of Humean supervenience, distinguishing a strong and weak version. §5 deals with the role of natural properties in the definition of duplicate entities and the principle of recombination, as required by Lewis's modal realism. In §6 we begin to look at the theory of constitution as a possible source of constraints by analyzing the admission of unlimited mereological complexity (the so-called *gunk*.) The discussion of gunk will be also the occasion to compare different characterizations of the "minimality" of bearers, in mereology and out of it. In §7 we study how composition as identity, as advocated by Lewis in *Parts of Classes*, fits with minimalism and its rival views. §8 draws some conclusions.

2. Minimalism *vs.* Liberalism

There are two alternative hypotheses about the bearers of natural properties. They are mutually exclusive, insofar as their definition makes clear that you cannot subscribe to both of them without contradiction, and they are

[5] The choice of the principles aims to give a reasonably adequate picture of Lewis's metaphysics. Moreover, these principles constitute quite a cohesive theoretical package and they are often jointly adopted by metaphysicians in the Lewisian tradition (e.g. Ted Sider, Daniel Nolan, Laurie Paul.) However, we can not analyze here the ties between these principles: as a result, the reader is free to assume that they are reciprocally independent, so that it is possible to drop one or more of them without being forced to drop the others too.

exhaustive because, if you think that natural properties are instantiated, you need to accept one of them. According to the first, *minimalism* about bearers (MB), all the bearers of natural properties are minimal entities.[6] On the other side, according to *liberalism* about bearers (LB), the bearers of natural properties are not exclusively entities of a minimal size. MB will be at the hearth of our discussion and we shall elaborate on its different facets in due course. We shall concentrate on LB here.

The central characteristic of LB is that it leaves unspecified how the bearers of natural properties should be identified thereby opening up some complex issues. In other words, minimal size is not a general criterion to identify the bearers of natural properties: does this mean that there is no general criterion for the identification at all, or is the criterion simply different from that of MB? If there is a criterion, we face two alternatives: a) there could be formal and general criteria to identify natural properties which do not involve size; b) there could be a size criterion not requiring that the bearers are minimal entities. Both appear to be unpalatable for different reasons.

The alternative a) is hard to implement: a general criterion not involving size can resort only to properties instantiated by the bearers that are abundant or conventional: this would indeed avoid circularity, in so far as abundant and conventional properties are not natural. However, it seems awkward that abundant or conventional properties identify the bearers of natural properties, as natural properties are expected to have some kind of explanatory priority over the former. Perhaps the criterion could make appeal to relations instead of properties: as according to MB the bearers of natural properties are at the bottom end in the net of relations of constitution, so LB would instead resort to another net of relations, e.g. the net of spatiotemporal relations. However, no intuitive reasons why the bearers of natural properties should be characterized by a distinctive spatiotemporal location (or by other positions in a net of relations) come to mind, leaving the burden of the proof to the supporter of this alternative.

[6] At this level we leave undecided the exact nature of *minimality*. MB is true if and only if all the bearers of natural properties instantiate one of the two features which – in a broadly Lewisian theoretical setting – can be seen as a kind of minimality: they are points or point-sized entities or they have no proper parts (and so they are mereological atoms). In §6 we will discuss the ties between these two characterizations of minimality.

On the other side, b) may be developed only in one way, since in mereology (due to the transitivity of parthood) there is only another privileged level beyond the minimal one: the maximal. As a result, b) could be developed as a form of *maximalism* according to which the bearers of natural properties are entities with a maximal size. This idea has been recently suggested by Jonathan Schaffer as a kind of *monism*, according to which there is actually only one maximal entity which is the bearer of natural properties: the universe.[7]

Given this picture, we can say that, once LB is adopted, the most plausible reaction is to deny the existence of a general criterion to identify the bearers of natural properties, at least until another criterion (such as the *maximalist/monist*) is provided and made independently plausible. If there is no general criterion, then the best thing to do could be to rely on natural sciences for the identification of the bearers of natural properties: because Lewis explicitly defers the identification of natural properties to physics, it is perhaps simply methodologically consistent to so defer the identification of their bearers.

This deference to science is in potential tension with some of the arguments we are going to provide in this paper against MB (e.g., in the discussion of *composition as identity* in §7.) In general, the strong conclusion in favor of LB is also a limitation to the scientific investigation about the bearers of natural properties: it excludes that the bearers are the minimal entities, even as a contingent matter of fact. Such outcome could be seen as incoherent with the motivations of LB and as a source of suspicion about the premises at play: in the context of an overall discussion of composition as identity (which falls beyond our purposes) this could be seen even as an argument[8] to reject composition as identity on the whole, or to reformulate it in order to avoid any necessary limitation on the size of the bearers of natural properties. After all, we are not assuming that all the theses we discuss, which could be traced back to a common Lewisian ground, should be accepted or rejected as a whole package.

Before moving over, we shall make explicit a delimitation of our problem. We have seen that Lewis distinguishes a restricted sub-domain of perfectly natural properties among natural properties. In this paper we will use the

[7] Cfr. (Schaffer, 2007) and, for a critical discussion, (Morganti, 2009).

[8] If the conclusion is unacceptable, composition as identity is not the only suspect premise. Another possibility is simply to reject or reform in depth the doctrine of natural properties.

expression "natural properties" for the sake of simplicity, but our arguments hinge on the perfectly natural properties.

It is unclear whether the distinction between two sorts of natural properties has consequences for the bearers of natural – but not perfectly natural – ones. Actually, it is not even set whether their bearers are different. According to a well-known suggestion by Lewis,[9] we can pass from the perfectly natural properties to the imperfectly natural simply by combining the perfectly natural with appropriate logical connectives, such as conjunction. If all the natural properties can be reached by chains of such simple logical combinations with properties of a same bearer, no new bearer gets involved and the bearers of natural properties are exactly the bearers of perfectly natural ones.

A para-syntactical conception of the degrees of naturalness, however, may be regarded only as an intuitive example, rather than as a full-fledged theory.[10] An alternative picture is that, when we have a complex entity whose features are determined by the perfectly natural properties of its components, then also the features of the complex entity inherit a certain degree of naturalness from its components. In this scenario, the domain of the bearers of natural properties would be different from that of the bearers of perfectly natural properties and its boundaries could even be vague, since no threshold of complexity would trace the boundary between the bearers of "minimally natural" properties from the bearers of definitely unnatural properties.

In conclusion, the concept of non-perfectly natural (but still natural) properties should be clarified in the context of the doctrine of natural properties. In the discussion to follow, we are going to assume only the core of the doctrine, focusing exclusively on perfectly natural properties.

3. Naturalness and Fundamentality of Properties

According to Lewis natural properties carve nature at its joints. This characterization is metaphorical and the metaphor is not transparent: it is not clear what nature exactly is, what are its joints and why these joints (whatever they are) should be characterized by a certain domain of properties. However, some expected features of natural properties can be easily inferred by the

[9] Cfr. OPW, p. 61 and (Lewis & Langton, 1998).
[10] (Sider, 1995) discusses this problem in depth.

metaphor or are explicitly declared by Lewis in his works. Let us analyze such features in order to see if they constrain in some way the size of the bearers.

A seemingly implicit feature is that natural properties are not conventional: they hold independently of their use in any kind of categorization operated by a cognitive subject. Still, it is not immediately clear how we should decide, for a specific property, if it enjoys the expected kind of independence and objectivity. Also, the bearers of natural properties could nonetheless be the subject-matter of a convention and, as a result, they will become the bearers of some conventional, non-natural properties. Let us suppose for example that the distinction between Europe and Asia is conventional. Europe and Asia will include as parts a certain number of electrons, whose charge and spin are assumed by Lewis as examples of perfectly natural properties. But, as a consequence of the conventional distinction between Europe and Asia, some electrons get the conventional property of being "European" and some others become "Asian": however the electrons are prototypical minimal entities, notwithstanding any convention concerning them.

Because some entities instantiate both natural and conventional properties, the distinction could be between those items which instantiate both natural and conventional properties and other ones instantiating only conventional properties. If MB were true, any non-minimal entity would belong to the second group. But it is not clear how to reverse the order of the reasoning and get an independent reason in favor of MB. Why the concept of naturalness as opposed to conventionality should imply that only minimal entities instantiate both kinds of properties? There is no such constraint, at least if we are not already committed to MB for independent reasons.

Two more features deserve to be analyzed. They are both made most clear in *New Work for a Theory of Universals,* Lewis's most elaborate text on the doctrine of natural properties (henceforth, NWU). Here he points to two main theoretical roles of natural properties concerning resemblance and causality respectively. We can try to see if these expected theoretical purposes require the bearers of natural properties to have a certain size.

First, according to Lewis, natural properties capture facts that are relevant for resemblance, and thus are the points of reference when we need to classify entities.[11] Facts of resemblance are pervasive: every kind of entity can be

[11] NWU, p. 13.

involved in a relation of resemblance;[12] natural properties capture the *relevant* facts of resemblance, while other less relevant kinds of resemblance could be captured – for example – by the property of being an European electron.

The relevance of facts of resemblance can help to discriminate between different properties, but it has no apparent consequence on the size of their bearers: why should the resemblances between electrons be more relevant than the resemblances between atoms or molecules? One could *assume* that the relevant facts of resemblance concern minimal entities, but such a move would bring only circular evidence in favor of MB.

The second pivotal theoretical function of natural properties is to capture the causal powers of things. The metaphor "carve reality at its joints" occurs sometimes in this context. Almost all properties are causally irrelevant, and there is nothing to make the relevant ones stand out from the crowd. Properties carve reality at the joints–and everywhere else as well.[13] While properties carve reality everywhere, natural properties carve it at its joints, which are – among other things – the causal links at the core of some scientific laws. This theoretical function of natural properties can constrain in some way the choice of their bearers: an information we might get is that the bearers are parts of reality or nature. As a result – for example – sets and numbers are perhaps[14] not good candidates for the role of bearers of natural properties. However, size is *prima facie* not involved: in which sense would an electron be more involved in causal relations than an atom, a molecule or even an organism?

The attribution of special causal powers to electrons and other minimal particles could be additionally fine-tuned: it should be admitted that some causal links involve also bigger entities, but these macroscopic causal links would be completely determined by the causal links involving their minimal parts. In this sense, the causal laws concerning atoms, molecules and

[12] Perhaps a minimal condition in order to be connected by relations of resemblance is to instantiate a property whatsoever (it does not matter if this property is natural), but this does not make size relevant.

[13] NWU, p. 13.

[14] Lewis would have been reluctant to classify them thoroughly as "abstract entities". See OPW, pp. 83–84, where Lewis claims that there is nothing wrong in the idea that a set is involved in a causal link (as in the common picture according to which a set of causes cause a certain effect). If this point of view is adopted, the role of naturalness in the theory of causality does not lead to the exclusion of sets from the domain of the bearers of natural properties.

organisms would be reducible to the causal laws concerning their minimal constituents, such as electrons. Even supposing that the idea is correct, it would not be a consequence of the mere concept of naturalness, but a substantive philosophical thesis. Lewis himself subscribed to a similar credo via the principle of Humean supervenience. We will discuss the principle in the next section, where we will see that its consequences for the size of the bearers of natural properties are rather weak, both in content and in modal force; but in any case they don't follow analytically from the expected features of natural properties.

The idea that minimal entities are endowed with some sort of primacy in the causal links of reality and that some of their properties account for these *fundamental* joints of nature lead us to the cognate notion of fundamental property. Fundamental properties are akin to natural properties, and there are several texts where Lewis seems to treat "fundamental" and "natural" as interchangeable attributes.[15] However, when a property is characterized as "fundamental", some considerations of economy are often involved.

The economy does not concern the single property but a class of properties providing an adequate grounding for something larger: properties are fundamental insofar as other properties (instantiated by other things or even by the same things) can be in some sense reduced to or made dependent upon them. Lewis's idea is roughly that non-natural, abundant properties can be reduced to natural ones, and in this sense naturalness and fundamentality are strictly connected. But the point of view of fundamentality involves the exclusion of those natural properties which, though non-conventional and relevant for natural laws, are not required in order to ground or explain a wider domain of properties: it is enough to consider a smaller domain of natural properties, leaving no explanatory roles for the others.

Fundamentality is a relative notion: a property is fundamental relative to a certain domain of properties, which should be grounded or explained by the fundamental ones. An example of such a domain could be the totality of properties instantiated by all the parts of a possible world *w*. In this case a property is fundamental in *w* if and only if it is natural and it is included in any basis upon which all the properties instantiated by all the parts of *w* are reduced.

[15] See the texts about Humean supervenience quoted in §1.

Some connotations in the concept of fundamentality lean towards MB. The *fundamentum* is easily identified with a minimal level of complexity, instantiating certain basic properties. The minimal level of complexity gives us the basis upon which all the features (and the causal links) of bigger entities *supervene*. But still, except for the lexical connotations of the term "fundamental", the entrenchment with minimality is a substantive philosophical thesis and not an analytic consequence of the concept of fundamentality, as we have defined it above relatively to a certain world. The intersection of any basis to which all the properties of every part of a world *w* can be reduced could include properties of entities of any size. The sizes could even be different in different worlds: for what follows from the mere definition of the concepts involved, the *fundamentum* could be given by atoms, electrons, molecules or organisms; the primacy of a certain level of complexity needs substantive arguments.

4. Humean Supervenience: Weak and Strong

Humean supervenience was seen by Lewis as the core of his entire philosophical work. According to the already quoted "Introduction" to PPII, Lewis actually got interested in some philosophical topics just in order to motivate Humean supervenience and defend it from some possible objections. In that passage, Humean supervenience is formulated so that perfectly natural and fundamental properties constitute the basis for supervenience and are said to be instantiated by points or point-sized entities:

> It is the doctrine that all there is to the world is a vast mosaic of local matters of particular fact, just one little thing and then another. […] Maybe points of spacetime itself, maybe point-sized bits of matter or aether or fields, maybe both.[16]

It should be immediately noted that this is not as decisive a declaration in favor of MB as it seems. After all, what can be inferred about the identification of the bearers of natural properties? Only that *some* "minimal entities" instantiate the natural properties which are in the basis of Humean supervenience. It does not follow that no non-minimal entity instantiates natural properties as well. It does not follow, unless one also assumes that the

[16] PPII, pp. ix–x.

only theoretical purpose of the doctrine of natural properties on the whole is to lay down a basis for Humean supervenience. This is a crucial point in the debate, one that we firmly resist: natural properties are key also in defining duplication and recombination; they play a role which is not instrumental to supervenience, rather it could be the other way round. Humean supervenience is compatible with MB, but it cannot require that only minimal entities instantiate natural properties.

The consequences are a bit stronger if we turn our attention from naturalness to fundamentality. In the quoted passage Lewis does not distinguish between perfect naturalness and fundamentality. But, if the distinction between them we have drawn in §3 is accepted, then it seems that the most relevant notion with respect to the supervenience debate is fundamentality: the obvious aim is to identify a most economical basis upon which everything else supervenes. The perfectly natural properties – as identified by scientific investigation or by the heterogeneous theoretical needs which the doctrine of natural properties is called to satisfy – could be redundant; on the other hand, as we have seen, the requirement of economy and non-redundancy is somehow inscribed in the notion of fundamentality itself. So, in a not very informative sense, Humean supervenience suggests that no non-minimal entity instantiates fundamental properties, because if a property is not in the minimal supervenience basis then it can be natural, but not fundamental.

The consequences of Humean supervenience for our problem are weaker than expected and problematic. It remains to see that their weight depends on the epistemological status and the modal force of Humean supervenience itself. The programmatic formulation of the introduction to the PPII conceals a deeper articulation; Humean supervenience was not meant by Lewis as a monolithic thesis[17] and it comes in two main versions: a weaker core, which is *a priori* and concerns every possible world, and a stronger thesis, which is *a posteriori* and concerns only our world and other worlds sufficiently similar to ours. The paper where this distinction is carried over most clearly is *Humean Supervenience Debugged* (henceforth, HSD). The following is Lewis's formulation of the weak core (here labeled WHS, Weak Humean Supervenience):

[17] Cfr. also (Nolan, 2005, pp. 28–29).

> If two possible worlds are discernible in any way at all, it must be because they differ in what things there are in them, or in how those things are. And "how things are" is fully given by the fundamental, perfectly natural, properties and relations that those things instantiate. (HSD, pp. 493–494.)

WHS is fully non-committal about the size or the complexity of the bearers of perfectly natural, fundamental properties (also in this case the two qualifications are treated as interchangeable.) With respect to the weak core, Humean supervenience in its strong version (SHS) is presented as "yet another speculative addition," concerning our world and "worlds like ours". The contents of this speculative addition are that: 1) the fundamental, perfectly natural properties are "local," in the sense that they are instantiated by points or point-sized entities; 2) the relations involved are spatiotemporal.

The constraints about the size of the bearers come from 1), thus they inherit their epistemological status and modal force from SHS. As a result, the partial and problematic evidence in favor of MB licenses MB, at best, as a contingent thesis.

5. The principle of recombination in modal realism

Natural properties have a very important role in Lewis's modal realism: they are called to make sure that for any possible way things might be there is a world where things are in that way. No genuine possibility should be passed over, otherwise, for example, our semantics risks licensing as necessarily true sentences which are only contingently so. This cardinal *desideratum* of modal realism is called *plenitude* in the first chapter[18] of OPW. Plenitude can not be simply stipulated, since worlds are expected to exist on their own, not as a consequence of a stipulation. Instead, plenitude needs to be grounded in an independently plausible metaphysical principle. Lewis thinks that this role can be played by the principle of recombination, according to which possible worlds are such that they respect our intuition that anything can coexist with anything and can fail to coexist with anything: according to this other broadly Humean intuition, there is no necessary coexistence between distinct entities. The totality of recombinations of distinct individuals should give us the expected plenitude of possible worlds.

[18] OPW, pp. 86–92.

Lewis's variety of modal realism cannot accommodate the intuition that anything can coexist with anything in the most straightforward way, that is admitting that any combination of pieces of possible worlds is itself a possible world.[19] Notoriously, Lewis's worlds do not overlap: no individual is in more than one possible world. As a result, the path to plenitude through the principle of recombination is a bit less direct and involves the admission of vicarious entities, called *duplicates*. A first, rough formulation of the principle could be the following:

> PRINCIPLE OF RECOMBINATION: Given any choice of parts of possible worlds, there is a possible world, which includes a duplicate of each part and nothing else.

A duplicate of an entity is an entity adequately similar to it. The expected kind of similarity is different from that involved in the counterpart relation under two aspects. First, the relation of duplication ought to be fully determined and exempt from any kind of vagueness, otherwise the domain of available recombinations would have vague boundaries and plenitude would not be definitely attained. Second, the properties in common between duplicate entities should not require the presence, in the same possible world, of the duplicate of something else. We have seen that an intuition to be respected in order to get plenitude is that anything can fail to coexist with anything: for this reason, duplicates should be allowed to differ in extrinsic properties, that is properties whose instantiation requires that there is a certain other entity in the same world. Thus, duplicates are required to share only intrinsic properties.

According to Lewis, the required kind of definite and intrinsic similarity can be obtained by stipulating that two entities are duplicates if and only if they have all the perfectly natural properties in common. This leads us back to our problem: any part of world which is recombinable according to the principle of recombination should instantiate at least one perfectly natural

[19] An adequate discussion of the principle of recombination should deal also with some constraints of size. Lewis was well aware of the importance of these constraints (OPW, pp. 90–92), which have also been discussed in the literature about the principle of recombination. Our formulation ignores this problem for the sake of simplicity.

property, otherwise it would be impossible to identify its duplicates.[20] Thus, any entity that can be recombined is a bearer of some natural properties.

The problem is that it is not clear which entities are recombinable. In the provisional, rough formulation above, we have involved every part of every possible world. But perhaps it is possible to obtain just the same domain of possible worlds and the expected plenitude recombining only the smallest pieces. Lewis presents the principle of recombination quite cursorily in a few pages of OPW and does not say what should be recombined. There is an open debate about the most economic, adequate formulation of the principle of recombination[21] and the prevailing opinion seems to be that it is not enough to recombine the smallest pieces. We can not review here this debate and we mention only the simplest reason to doubt that atoms are enough: if there are worlds with no atoms (call them *gunkish worlds*) or where some parts of the world are not composed of atoms, then it is not clear how the principle of recombination should be applied to these worlds. If all or some of these *gunkish* worlds are not (vicariously) recombinable, then plenitude is unattained.

Anyway, it seems that no outcome of the debate about the principle of recombination would be really favorable to MB. If the principle of recombination needs to involve also bigger or more complex entities, then an important aspect of modal realism implies LB, since non-minimal entities need to have duplicates and thus to instantiate perfectly natural properties. If instead the atomistic formulation of the principle of recombination can be made plausible through some adjustments, then any atomic part of any possible world will instantiate at least one perfectly natural property. But, even in this scenario – which is seemingly unsympathetic with LB – we could not conclude that nothing else instantiates perfectly natural properties. As in the case of SHS in §4, the consequences are at most positive, but not negative: since perfectly natural properties are not introduced for the sole purposes of the principle of recombination, the principle can require that something instantiate natural properties, but can not exclude that something else instantiate them too.

[20] Two entities which do not instantiate any perfectly natural properties have trivially in common all their natural properties. This trivialization should be avoided in a proper definition of duplication, which is beyond our purposes in this paper.

[21] See in particular (Efird & Stoneham, 2008) and (Darby & Watson, 2010).

Moreover, the principle of recombination is a problematic aspect of modal realism, and has been criticized under several points of view.[22] This discussion could make MB even less plausible. For example, let us consider a minimal part of our world: an electron. A duplicate of an electron in a given different world should share with it all its perfectly natural properties, such as its spin and charge. Does this guarantee that the duplicate of the electron is a *minimal* part of the given world? Why should the sharing of charge and spin imply that the duplicate of the electron has a point-like spatial extension or no proper part? If a minimal entity has a non-minimal entity as one of its duplicates, we get a non-minimal bearer of natural properties. This problem deserves closer attention than the one we can give it here. Nevertheless, a joint supporter of the principle of recombination and of MB has the burden of explaining why there is no relation of duplication of this kind.

6. Mereology and Gunk

We have seen that the worlds or parts of worlds which are not composed of atoms are problematic cases for the combination between MB and the principle of recombination in modal realism. The so-called *gunk* was admitted by Lewis as a genuine possibility. The admission of this possibility was deeply connected with Lewis's idea that classical mereology is just the general, exhaustive theory of ontological constitution. It is exhaustive not only in the sense that the hypothesis of a non-mereological kind of constitution is not tenable,[23] but also in the sense that all the kinds of constitution which are licensed by classical mereology are genuine possible ways in which a world might be. Classical mereology is not committed to atomicity, thus it is a genuine possibility that some entities (or even the world in its entirety) are not composed of atoms, so

[22] The following are some other problematic aspects of the principle of recombination: is the intuitive principle that anything can coexist with anything sufficient to guarantee plenitude? Is it enough to recombine pieces of possible worlds directly or should we require recombinations of properties themselves? Is the distinction between intrinsic and extrinsic properties really sharp? Is the sharing of perfectly natural properties a sufficient condition for the sharing of intrinsic properties and does it really allow for a free variation of extrinsic properties? We thank John Divers for the suggestion that an adequate formulation of the principle of recombination (one that provides genuine plenitude) could lead to an open rejection of MB, if not to the rejection of the doctrine of natural properties on the whole.

[23] Cfr. (Lewis, 1992).

that their proper parts have always still further proper parts. Via *plenitude*, it follows that there are worlds entirely or partially non-atomic.

In this section we argue that the incompatibility between the admission of *gunk* and MB is more general, while it is possible to combine *gunk* and LB in different ways. However, a preliminary clarification is needed. The admission of *gunk* can be easily seen as the rejection of minimal entities: if anything has still further parts, nothing is really minimal since there are always smaller things. However, the concept of minimality here at work is *prima facie* different from the one employed for example in the strong formulation of Humean supervenience, where the fundamental properties in the basis of supervenience were said to be instantiated by points or point-sized elements. In the case of SHS, the typical minimal entities are points.

What is a point? The question is difficult and, as far as we know, Lewis has never taken side or expressed an opinion in print about it. However, in the passages about SHS quoted above, the properties in the supervenience basis are said to be local because they are instantiated by points or point-sized entities. This suggests that points get involved insofar as they have a minimal extension (a minimal localization) in space and time.

Is it legitimate to identify points with mereological atoms? The answer to this question is pivotal for us. Indeed, in this and the following sections we are going to draw some conclusions from two mereological principles – the admission of *gunk* and the so-called thesis of composition as identity respectively. However, if the mereological characterization of minimality were completely extraneous to that presupposed in some important aspects of the doctrine of natural properties (such as the discussion of SHS), the consequences of the mereological principles could not interact with the outcome of our analysis in the previous sections.

And actually it is easy to point to examples of mereological atoms that are not points, even if the examples are unavoidably relative to one's ontological commitments. In Lewis's *Parts of Classes* for example, set-theoretical singletons have no proper parts, thus they are atoms: but the singleton of the number 0 has no spatiotemporal extension at all, because it is not an entity in the spatiotemporal domain. In some kinds of theories of universals, universals

are parts of the individuals instantiating the corresponding property.[24] Moreover some universals are simples (as they are not composed of other simpler universals.) These universals will be atoms, but the theory is still free to deny that they have any spatiotemporal extension at all: they are not points or point-sized entities in any sense. The theory of universals can provide also an example of a point-sized entity which is not a mereological atom: an electron with a negative charge (as we have seen, one of Lewis's preferred examples of minimal entity) would have the universal of negative charge as one of its parts; as a result it would not be a mereological atom.

However, it is easy to restrict mereological atomicity to the spatiotemporal domain. In general, a mereological atom is an entity which has no proper part. A spatiotemporal mereological atom will be an entity with a spatiotemporal extension which has no proper part with a spatiotemporal extension. In this sense the electron is a spatiotemporal mereological atom, while the singleton of the number 0 and the universal of negative charge are not spatiotemporal mereological atoms, since they have no spatiotemporal extension.

It is worth remarking that this restriction of mereological atomicity does not impair the validity of the principles of the mereological theory of constitution which we are going to review. The admission of *gunk* has no peculiar connection with sets or universals, and the most intuitive example of gunk is probably given by the indefinite divisibility of space. As for the thesis of composition as identity – which we are going to review in the next section – it concerns any kind of composition, including the most obvious cases of spatiotemporal parthood. It is thus legitimate to draw conclusions from these two mereological principles on the minimal or non-minimal size of the bearers of natural properties, with the *proviso* that the mereological minimality which is at play is not general mereological atomicity, but restricted spatiotemporal atomicity.

We can now proceed to evaluate the consequences of the admission of gunk on the size of the bearers of natural properties. If gunkish worlds (or parts of worlds) are admitted, the problems for MB are not limited to the best

[24] See for example (Armstrong, 1978), although it should be remarked that he changed mind on this point in later versions of his theory of universals. An analogous example could be easily built with tropes, since according to many trope-theorists – including the classic (Williams, 1953) – ordinary individuals are mereological sums of tropes.

formulation of the principle of recombination which we have discussed in §5. MB implies that no natural property is instantiated in a gunkish world or part of world. Yet, the lack of natural properties brings some problematic consequences with itself: gunk would be, so to say, inert – excluded from the domain of possible recombination, not relevantly similar to anything, devoid of causal powers. This scenario, if not provided with an independent motivation, seems unacceptable.

By contrast, LB can cope with *gunkish* worlds and parts of worlds in two general ways. There could be: 1) either an infinite descent of bearers of perfectly natural properties; 2) or a privileged level under which no natural property is instantiated. In a passage of *Against Structural Universals,* Lewis discusses briefly 1):

> I note that class nominalism, with a primitive distinction between natural and unnatural classes, has no problem with infinite complexity. It might happen that whenever we have a natural class, its members are composite individuals, and their parts (and pairs, triples... of their parts) fall in turn into natural classes. (OPW, pp. 86–87)[25]

Here we can not discuss in depth the two options. We note only that both, when given an adequate articulation, are likely to provide a criterion for the identification of the bearers of natural properties which will be applicable also to non-gunkish worlds and parts of worlds. Once this criterion is conceded for the special case of gunk, the restriction of the criterion only to gunk seems arbitrary. The quotation above shows that Lewis concedes 1) for gunkish worlds or parts of worlds: how could he deny in a principled way that, even in fully atomic worlds, the bearers of natural properties are distributed at several levels of mereological complexity? On the other hand, the kind of criterion invoked in 2) will probably not be mereological, since, due to the transitivity of the relation of parthood in classical mereology, there is no mereologically

[25] The reference to class nominalism does not mean that 1) is compatible only with a certain stance on the problem of universals, namely with Lewis's own class nominalism. In the context of the quotation, Lewis is criticizing the reasons leading David Armstrong to admit structural universals. Lewis has already conceded that the theory of structural universals is able to cope satisfyingly with infinite complexity, but remarks here that class nominalism has no problem too. In the immediate following he observes also that "likewise a trope theory has no problem with infinite complexity". We quote the passage about class nominalism because it includes the most explicit admission of an infinite descent of bearers of natural properties.

privileged level, except the (eventual) level of atoms and the maximal level of the universe.[26] But if the invoked criterion is non-mereological (if it involves, for example, some kind of unity or cohesion), it is possible to wonder why should it be applied exclusively to gunkish worlds and parts of worlds. If the criterion works in this case, a specific, independent motivation for the restriction to these cases should be provided. In absence of such a motivation, LB can be indefinitely extended from gunkish to non-gunkish scenarios.

7. Composition as Identity and Boring Composition

According to a pivotal thesis of Lewis's *Parts of Classes* (henceforth, POC), composition is a kind of identity.[27] For this reason the thesis is usually labeled as CAI (composition as identity). The analogy between composition and strict one-one identity holds allegedly under several respects,[28] but only one of these points of resemblance is relevant for our purposes. It is the so-called *ease of description*,[29] according to which, once you have described exhaustively some entities, no further effort is required in order to describe exhaustively their sum. Conversely, an adequate description of a whole gives also an adequate description of its parts. Something analogous happens with one-one identity: when you describe an entity *x*, you describe *ipso facto* also everything which is identical to *x*.

Ease of description for one-one identity is a trivial consequence of the principle of indiscernibility of identicals, according to which identicals share all their properties; if an exhaustive description captures all the properties of an entity, no other property needs to be captured for those identical to it. However, in the case of composition, the principle of indiscernibility cannot hold, since it is very easy to point at properties instantiated by the whole but not by its parts (or *viceversa*): for example, as Lewis himself remarks, a piece of land is one, while the six parcels composing it are six.[30] Some properties can well be common to whole and parts (for example, both the parcels and the bigger piece are pieces of land), but in general the different ways of

[26] We have briefly discussed the maximalist-monist alternative in §2.
[27] Cfr. POC, pp. 81–87.
[28] See the paper by Carrara and Martino and the commentary on POC by Bohn in this volume for an overall analysis of CAI.
[29] POC, p. 85.
[30] POC, p. 87.

partitioning the same stuff bring with themselves many important, not only numerical properties. For example, let us suppose that the piece of land is rectangular while the six parcels are square.[31] Obviously, the rectangular piece of land is also the fusion of the two triangular pieces obtained tracing a diagonal of the rectangle. The two triangles, the six squares and the big rectangular piece of land differ not only in number, but also in the fact that in the first case we have triangles, in the second squares and in the third a rectangle.

How does ease of description work for composition, if indiscernibility does not hold in this case? Both the six squares and the two triangles of land *compose* the big piece of land. Thus, Lewis's thesis is that, for example, an exhaustive description of the six squares of land gives also an exhaustive description of the big rectangular piece and an exhaustive description of the two triangles; and the same happens if we begin with a description of any other partition. Lewis, in the few pages of *Parts of Classes* devoted to CAI, mentions explicitly some exceptions to indiscernibility between whole and parts, as a reason to restrict the analogy between composition and one-one identity: thus, he was clear that ease of description for composition does not rely on indiscernibility, but he does not say how it works instead. A plausible interpretation, quite consonant with Lewis's philosophy in general, is that ease of description is connected with supervenience: an adequate description of a certain partition of some stuff specifies the properties of that stuff partitioned in that way and, if there is more than one piece, the relations between different pieces. These properties and relations are not, in general, common to the other partitions of the same stuff; instead they *determine* the properties and the relations for the other partitions; thus, they determine what needs to be captured by an adequate description for them. Thus, for example, a description of the big rectangle of land gives also a description of the six squares not because it specifies also all their properties, but because it specifies the properties of the rectangle, which on their turn determine the properties of the square and the relations between them.

[31] This example is partially borrowed from the paper by Carrara and Martino in this volume (see their section 4, where – however – the composition is directly between geometrical items instead of pieces of land), but we interpret it in a different way.

According to an efficacious expression of Jonathan Schaffer,[32] composition as identity makes composition *boring*. When you go up and down through different levels of size and complexity in the mereological structure of reality, you never incur in surprises. An adequate description operated at a certain level of complexity and size is also an adequate description at any other (more or less fine-grained) level of complexity and size. This happens because the relevant features are co-determined: just specify them at a level whatsoever and those at any other level are immediately determined. In this light, we can define boring composition in terms of supervenience: composition is boring if and only if the properties of the whole supervene on the properties of the parts and the relations between them *as much as* the properties of the parts and the relations between them supervene on the properties of the whole.

Natural properties are obviously among the features captured by an adequate description: even if Lewis does not explain what exactly counts as an adequate description, it is reasonable to expect it to specify natural properties, which – as we know from §3 – account for relevant similarities and causal links. Now, CAI is not seen by Lewis as a restricted phenomenon: every composition is a kind of identity and is analogous to one-one identity also for what concerns ease of description. As a result, every case of composition comes out as boring. But if composition is always boring, on what basis should the properties of a certain level of size or complexity be deemed more natural? Composition as identity confirms what SHS, as we have seen in §4, literally says: all the relevant features of things of any size supervene on the relevant features of entities at the simplest, atomic level. But it adds something else, utterly incompatible with the idea that there is a fundamental level endowed with a sort of objective primacy: the relevant features of entities at the simplest atomic level supervene on the relevant features of entities at any other level.

When composition is boring and supervenience is symmetric, the properties of the compound are as natural and as fundamental as the properties of the components. Perhaps there could be still some epistemological strategy to advocate the idea that the properties at the atomic level enjoy some kind of explanatory primacy. Although according to composition as identity any level guarantees an adequate, exhaustive description, it could be epistemically convenient to provide a general criterion about the level at which we should

[32] Cfr. (Schaffer, 2003, p. 505).

start, and perhaps the atomic level is the easiest to identify. Still, there is no metaphysical motivation for this alleged primacy.

Composition as identity is incompatible with MB. Moreover it leads to an extreme version of LB. It is not that natural properties are spread sparsely at different levels, instantiated for example by unified bodies or organisms of some kind: natural properties are simply everywhere, instantiated at any level of complexity. This consequence could be in contrast with some pristine motivations of the supporters of LB. Perhaps, they wanted to show that some important, objective, irreducible properties are instantiated by complex, structured entities: *emergent* properties. These properties are emergent, by definition, insofar as they do not supervene on the properties of the constituents and the relations between them. Such kind of *emergentism* is likely to reject composition as identity and embrace LB; according to it, emergent properties – such as acidity and proprio-ception – are just irreducible natural properties instantiated by non-atomic entities. By contrast, composition as identity is a thesis typically endorsed by those – like Lewis – who are reluctant to admit emergent properties. The same idea of *boring composition* is the utter denial of emergentism: the surprises you do not incur in are just emergent properties.

It is interesting to remark that CAI and emergentism – while being two so radically different and exclusive views – share at least a consequence when conjoined with the doctrine of natural properties: the rejection of MB. For what concerns the size of the bearers, they differ only in the flavor of LB they license: while emergentism tends to attribute natural properties to unified or cohesive bodies, CAI scatters natural properties at every level of the compositional structure of reality.

8. Conclusion: Are MB and SHS Necessarily False?

The survey of the main metaphysical tenets in Lewis's metaphysics that are relevant to the question of the bearers of natural properties suggests several considerations. As for supervenience, SHS implies only that, in "worlds like ours" some point-sized entities are the bearers of natural properties; WHS instead is non committal about the bearers of natural properties. As for recombination, we have seen that the principle governing it does not clearly suggest that only atoms should be recombined and does not exclude that non-minimal entities are bearers of natural properties. On the other hand, the

admission of gunk provides a case where the restriction of natural properties to atomic bearers cannot happen *and* where any alternative to this restriction seems applicable also to non-gunkish situations. At the same time, CAI suggests that composition is "boring;" as a result, it is not clear why the properties of the whole should be less natural than the properties of its parts.

In this general picture, the destinies of MB and SHS are intertwined: if WHS is true, then MB is true only if SHS is true; if WHS holds and SHS fails, it is because there are non-atomic bearers of natural properties. Now, SHS is *at best* contingent: it does not follow from the theory of constitution, which holds necessarily. One possible move, adopted by Lewis and more recently endorsed by other authors[33] is to hold that SHS is true of the actual world. But this seems questionable for three reasons: 1) for all we know, the actual world may be gunkish (we have no principled way of ruling out this possibility); 2) if the restriction of natural properties to atomic bearers is waived for gunkish worlds, not to waive it for all worlds appears as calling for further justification; 3) boring composition (implied by CAI, which is meant as a necessary principle) rules against SHS.

On the score of these results, we conclude that nothing in Lewis's metaphysics justifies the conclusion that, *necessarily*, only atoms are the bearers of natural properties. Lewis's mereological theory of constitution suggests that natural properties are instantiated by entities of any level of complexity in most worlds; we have no principled way of telling whether our world is one of those. In Lewis's metaphysics, there is a tension between CAI (boring composition) and SHS, even when the latter is regarded as a contingent claim. Unless the tension is resolved, both MB and SHS risk coming out as necessarily false.

[33] See for example (Nolan, 2005, pp. 28–29) and the supplement "The Contingency of Humean Supervenience" in (Hall 2010).

REFERENCES

Armstrong, D. (1978). *Universals and Scientific Realism*, Cambridge: Cambridge University Press.

Darby, G. & Watson, D. (2010). Lewis's Principle of Recombination: Reply to Efird and Stoneham. *Dialectica*, *64* (3), 435–445.

Efird, D. & Stoneham, T. (2010). What is the Principle of Recombination? *Dialectica*, *62* (4), 483–494.

Hall, N. (2010). David Lewis's Metaphysics. The Stanford Encyclopedia of Philosophy (Fall 2010 Edition), E. N. Zalta (ed.). URL = http://plato.stanford.edu/archives/fall2010/entries/lewis-metaphysics.

Hüttemann, A. (2004). *What's Wrong with Microphysicalism*. London: Routledge.

Karakostas, V. (2009). Humean Supervenience in the Light of Contemporary Science. *Metaphysica*, *10*, 1–26.

Langton, R. & Lewis, D. (1998). Defining 'Intrinsic'. *Philosophy and Phenomenological Research*, *58* (2), 333–345, reprinted in Lewis, D. (1999). *Papers in Metaphysics and Epistemology*, Cambridge: Cambridge University Press, 116–132.

Lewis, D. (1983). New Work for a Theory of Universals. *Australasian Journal of Philosophy, 61*(4), 343–377. (NWU)

Lewis, D. (1986a). *Philosophical Papers II*. Oxford: Blackwell. (PPII)

Id. (1986b). *On the Plurality of Worlds*. Oxford: Blackwell. (OPW)

Id. (1986c). Against Structural Universals. *Australasian Journal of Philosophy*, *64*(1), 25–46, reprinted in Lewis, D. (1999). *Papers in Metaphysics and Epistemology*, Cambridge: Cambridge University Press, 78–107.

Id. (1991). *Parts of Classes*. Oxford: Blackwell. (POC)

Id. (1992). Armstrong on Combinatorial Possibility. *Australasian Journal of Philosophy*, *70*, 211–224; reprinted in Lewis, D. (1999). *Papers in Metaphysics and Epistemology*, Cambridge: Cambridge University Press, 196–214.

Id. (1994). Humean Supervenience Debugged. *Mind*, *103* (412), 473–490. (HSD)

Morganti, M. (2009). Ontological Priority, Fundamentality and Monism. *Dialectica*, *63*(3), 271–288.

Nolan, N. (2005). *David Lewis*. Durham: Acumen.

Schaffer, J. (2003). Is there a Fundamental Level? *Noûs*, *37*(3), 498–517.

Id. (2004). Two Conceptions of Sparse Properties. *Pacific Philosophical Quarterly*, *85*(1), 92–102.

Id. (2007). From Nihilism to Monism. *Australasian Journal of Philosophy* , *85*, 175–191.

Sider, S. (1995). Sparseness, Immanence and Naturalness. *Noûs*, *29*, 360–377.

Williams, D. (1953). On the Elements of Being I. *Review of Metaphysics*, *7*, 3–18.

The Modal Dimension*

Giuliano Torrengo†
giuliano.torrengo@ub.edu

ABSTRACT

Space and time are two obvious candidates as dimensions of reality. Yet, are they the only two dimensions of reality? Famously, David Lewis maintained the doctrine of "modal realism", the thesis that possible worlds exist and are entities as concrete as the actual world that we live in. In this paper, I will explore the idea that modality can be construed as a dimension along with space and time. However, although Lewis' modal realism is the main source of inspiration for this construal of modality, I will argue that something else is required for having a modal dimension.

1. The Many Dimensions of Reality

Space and time are often thought of as two *dimensions* in the "arena" that contains all *material entities*.[1] Material entities can be thought of as existing in reality by being located in both the temporal dimension and the spatial dimension. If an entity exists in a certain dimension *d*, then it has (proper or improper) *d*-parts *located* in or along *d*. This notion of dimension – to be made more precise in what follows – is not to be confused with the notion of an

* I thank the editors, the participants at the Seminar "Parts and Constituents in Analytic Metaphysics" at the Scuola Normale in Pisa (10 December 2010), the participants at the PERSP Metaphysical Seminar at the University of Barcelona (18 March 2011) for their useful comments, and I acknowledge the financial support of the project FFI2011-29560-C02-01 of the MICINN.

† LOGOS – University of Barcelona, Spain. Labont – University of Torino, Italy.

[1] All and only? Space and time may be inhabited by physical entities that are not material (e.g. fields, forces, powers, vectors, tropes). Besides, there may be immaterial entities existing in time (e.g. souls). In what follows, I shall not consider immaterial entities and non-material physical entities, but I will not rely on the "and only" assumption either. Roughly, I assume that having causal efficacy is a necessary but not sufficient condition for being material and that possessing ordinary causal efficacy is sufficient for not being immaterial.

entity *having* a spatial dimension or a temporal dimension. For a material entity to have a spatial dimension or a temporal dimension in a non-trivial sense it must have spatial *proper parts* or temporal *proper parts* respectively (any entity that exists in space has a spatial dimension in a trivial sense, since it has an improper spatial part – viz. itself – and *mutatis mutandis* for other dimensions). However, if an entity exists in a dimension *d*, it does not follow (at least not analytically) that it possesses a *d*-dimension in a non-trivial sense – viz. that it has proper *d*-parts. For instance, an entity may exist in the temporal dimension but lack proper temporal parts because point-like (e.g. instantaneous events), or because it is multi-located at different points in the dimension (e.g. enduring objects), or because it is temporally extended but simple.

Are space and time the only two dimensions of reality? Roughly, to take seriously the talk of a dimension *d* is to consider the *fact that d is a dimension* as a fundamental fact – something that cannot be explained by means of anything else. Contrariwise, to deny that *d* is a dimension of reality is to consider talk of *d* as a dimension as derivative on other facts (for instance, although this is not the only option, on facts concerning the way we conceptualise things).[2] Famously, David Lewis maintained the doctrine of "modal realism", the thesis that possible worlds exist and are entities as concrete as the actual world that we live in. In what follows, I will explore the idea that modality can be construed as a dimension along with space and time. In fact, although Lewis' modal realism is the main source of inspiration for this construal of modality, I will argue that something else is required for having a modal dimension.

[2] I am using a sense of the derivative/fundamental pair that implies some form of anti-realism with respect to what is derivative (and, as we shall see, there are at least two construals of it). Of course there are other interesting senses of derivative/fundamental (e.g. composite and structured vs. simple and unstructured, less natural vs. more natural) that do not have anti-realist connotations, but these need not concern us here. See also (Williams, 2010). Moreover, the issue of the reality of a dimension is distinct from the traditional debate of the relationism vs. substantivism of space and time (or spacetime). In particular, space, time and modality may all turn out to be dimensions of reality in the same sense, even if it is not the case that for all of them either relationism or substantivism only is true.

2. Dimensions, Points, Entities

The arena of reality is a 'space', in the mathematical sense of an order of elements along a certain number of 'dimensions'. A dimension of reality can be a grouping of more 'dimensions' in this mathematical sense. For instance, we usually consider space to be a dimension that has three 'dimensions', while time is a dimension that has only one. I call the characteristic ordered elements of a dimension *d* the points of *d*. Generally speaking, the material entities that occupy a dimension *d* exist or are located at its points. In a sense, along a dimension *d* we find the points of *d*, but in another sense we find the entities that are located at them. How material entities behave with respect to *d* will depend on the specific features of *d* itself. For instance, if points of the spatial dimension are unextended, then we cannot find two distinct proper parts of the same thing at the same point (if they are extended than they are "regions", and of course the same object can have different spatial parts in the same region of space).

A central question to ask with respect to each dimension is whether the *same* entity can exist at more than one point along that dimension. There are two senses in which the same entity can be at more than one point along a dimension. The first one is the *strict* identity sense, and I take it to be primitive: at two different points of the dimension we find the *very same* entity. The second sense is the *loose* identity sense; it is a kind of similarity, that is, it does not imply strict identity, but only the sharing of a relevant property. The property that the entities must share in order to be identical in this loose sense is *being part of the same whole* (where "sameness" is identity in the strict sense). Hence loose identity can be defined in term of the part-whole relation (plus strict identity, and existing / being located at a point):

> (LI) The same (in the loose sense) entity is located at more than one point of *d* if and only if different parts of it are located at those points.

Typically, the spatial dimension is such that a material entity can be at more than one spatial point only in the loose sense, i.e. only by having distinct parts at different spatial points. I take this "Lockean" principle to be on the whole quite unproblematic.[3] The temporal dimension, to the contrary, is usually

[3] I am taking the principle for granted only because the spatial dimension is not the main focus of the paper: I am using it only as a clear case. However, at a macro-level of material objects at least the

taken to be such that the very same entity can be at different temporal points (viz. instants) in the strict sense (and at different temporal points it can be at different spatial points or it can "come back" at the same spatial point) – or at least, that is the construal of it that the so-called "three-dimensionalists" give. "Four-dimensionalists", in contrast, think that the restriction on the spatial dimension applies to the temporal dimension too, and that an entity can be at different instants only in the loose sense, or at least that physical entities can have different parts at different times also along the temporal dimension (and this is how entities persist in time) (Lewis, 1986; Heller, 1999; Sider, 2001; Hawley, 2001). Let us generalise the talk of "spatial parts" and "temporal parts" to other dimensions, as follows: for any dimension *d* and entity *x* that exist in *d*, call the "*d*-parts" of *x* those proper parts of *x* that exist at points of *d*, compose *x*, and are specific to *d* (hence *s*-parts are spatial parts and *t*-parts are temporal parts).

Along with the two senses of being the same at different point of a dimension *d*, I shall define the notion of "entirely existing at a point *p*" in terms of existing / being located at a point *p* of *d* and mereological notions:

> (EE) An entity entirely-exists at a point *p* of dimension *d* (or it is entirely located at *p*) if and only if all its *d*-parts exist at *p*.

Existing or being located at a point *p* does not imply entirely-existing or being entirely located at *p*. With respect to the spatial dimension, it may be the case that a spatially extended whole exists at a spatial point *p*, even if it is "larger" than *p* – namely even if it is not entirely located at *p*. With respect to the temporal dimension, again, we find philosophical disagreement. Three-dimensionalists maintain that, at least with respect to the kind of entities that are material *objects*, if an object exists at a certain instant *t*, then it also entirely exists at *t* (since the only temporal part that it has is itself). Four-dimensionalists maintain that there are many material objects that have

principle is appealing. It may be questioned whether micro physical entities at the quantum level still obey this restriction (French, 2011), or whether it holds for bundle theories of individuals (Rodriguez-Pereira, 2004). This is not to say that alleged counterexamples, even at the macro-level, are not discussed in the literature: see Fine (2000). A related though distinct issue is the possibility of co-location.

temporal parts, and thus they fail to entirely exist at the points at which they exist.[4]

Now, the spatial dimension and the temporal dimension "intersect" in the sense that the same (at least in the loose sense) entity that exists at a point *l* of the spatial dimension can also exist at a point *i* of the temporal dimension (and vice versa) – although the details will depend on the constraints that different metaphysical theories put on the relation between entities and points. This holds for any dimensions of reality in general – the idea behind being that dimensions have to be somehow connected to each other to be dimensions of the *same* reality. But what about further dimensions of reality beside space and time? Can there be any?

3. Can Modality Be a Dimension?

To anyone acquainted with modal logic, it is obvious that modality can be thought of as a dimension of reality whose points are *possible worlds*. What is less clear is whether the concept of modality is that of a dimension only in a metaphorical sense, in that its similarities with time and space are merely superficial, or whether there is a cogent sense in which it can be thought of as dimension of reality alongside time and space. Any framework *d* in which we talk of points of a dimension and entities existing at them, or being located at them, represents a dimension of reality in a *conceptually cogent* sense (and not in a merely *metaphorical* sense) if and only if it satisfies the following conditions:

(I) material entities exist / are located at least at some points of *d*;
(II) *d* intersects with other dimensions.

If modality is a dimension, according to (I), we should possibly find material entities located at different points of it. As we will see in the next paragraph, the two main metaphysical theories about modality disagree on how to construe the notion of existing or being located in a possible world. However, it is clear that if we cannot make sense of the idea of having material entities

[4] I am simplifying since I do not need to go into the details. In the literature, the notions of being entirely located at and being precisely located at a spatio(temporal) region are often taken as more primitive (e.g. Casati and Varzi, 1994; Calosi 2010). Besides, I am not considering four-dimensional relativistic spacetimes (Balashov, 2010), because they are immaterial to my main point.

located in possible worlds the whole business of modality would not be of much worth. As for (II), it appears more problematic. Often, possible worlds are thought of as containing entities that are connected to each other by spatial and temporal relations, but which are not connected by *any* spatial or temporal relation with things in other possible worlds. In other words, I am *not* very, very far in space and time from any of the things that could have existed; rather, there are no spatial or temporal relations whatsoever between things that exist in distinct possible worlds. However, as we noted for (I), if we could not make sense of the idea of a material object being at the same point of time and space but in a different possible world, the whole construal of modality in terms of possible worlds would be useless. Again, different metaphysics will be likely to give different construals of the intersection, but modality on the whole can be a dimension in a conceptually cogent sense.[5]

4. Is Modality a Dimension?

Once we have granted that a certain framework represents a dimension in a conceptually cogent sense, we should then ask what it takes for the represented dimension to be real. This is the working hypothesis: a dimension *d* is real if and only if both of the following obtain

(i) all its points are real in the same sense, and
(ii) its points ground truths concerning attribution of the relevant "dimensional" properties (to be specified below).

Let us first concentrate on (i). As we said in Section 1, realism with respect to a dimension *d* is tantamount to taking *d* to be a fundamental, not reducible, feature of reality. However, that does *not* imply that the points of *d* must be real in some irreducible sense too. Typically, metaphysical theories of modality disagree on what possible worlds are. David Lewis argued at length that all possible worlds are *concrete*, *material* entities composed of many other material entities as their parts (Lewis, 1986). Many philosophers disagree with Lewis and argue that there is an important difference between the world that we inhabit – the actual world – and the merely possible worlds. According to the so-called "actualist", the actual world is material and concrete, but possible

[5] The question whether a purported "dimension" should be an order in some precise sense to be a dimension in a proper sense will not be addressed here. See (Over, 1986; Lowe 1986).

worlds other than the actual one are abstract entities – usually sets of propositions or sets of sentences. Actualists do not deny that possible worlds exist, but claim that they are constituted by elements of the actual world, such as primitive modal properties, or sentences that represent how things might have been. Thus, actualism fails to meet (i), because possible worlds other than the actual one existentially depend on the actual world and are not on a par with it. For the actualist, modality is at best a limiting case of a dimension, namely a dimension containing only *one* point – the actual world.

The actualist also fails to satisfy condition (ii) for realism with respect to modal dimension. The idea behind (ii) is that if a dimension *d* is real, the points of *d* will play a crucial role in explaining why attributions of dimensional properties to (elements of) reality turn out to be – in the right circumstances – true. In order to outline this notion of "dimensional properties", it is important to distinguish between three kinds of properties that entities can have with respect to a dimension.

There are *simple properties* with respect to a dimension *d*. I take this to be a primitive notion. Roughly, a simple property P with respect to *d* does not make any "reference" to points of *d*, and whether an entity *x* has or does not have P is something that can be settled only relative to points of *d*. For instance, *being red* can be seen as a simple property with respect to time (incidentally, also with respect to space and modality), because there is nothing in *being red* that "points" toward one instant or other, and something may be red at certain times and not red at others.

A *d-indexed-property* P-*p* is constituted by a simple property P and a point *p* of a dimension *d*. The constituent *p* of P-*p* determines the conditions for possessing P-*p*. More precisely, an entity *x* possesses a *d*-indexed-property P-*p* if and only if *x* has P relative to *p*. Possession of a *d*-indexed-property is not something that varies across the dimension. For instance, consider the *t*-indexed-property being *red-at-*t_0. Something is red-at-t_0 if and only if it is red (simple property) at t_0, and the fact that something has or does not have such a property does not vary through time.

A *d-tensed-property* P-*tens* is constituted by a simple property P and a tensed element *tens*. As for simple properties, *d*-tensed-properties are possessed or not only relative to points of *d*. But as with *d*-indexed-properties, conditions for possessing them involve reference to points of the dimension. However, such a reference is only contextually determined by the tensed element *tens* of the property. More specifically, an entity *x* has P-*tens* relative

to a point *p* if and only if *x* has P relative to one or more points *p'* of *d*, which are related to *p* in a certain way determined by *tens*. The most common type of *d*-tensed properties are the *t*-tensed properties (aka "tensed properties" without further qualifications). An example of a tensed property is *having been green*. Let us suppose that there is a red apple in front of us. To this apple we can attribute *now* the property of having been green, and we would be right if and only if in the past the apple had had the simple property *being green*.[6]

d-tensed-properties are the "dimensional" properties we were looking for. They can, and usually do, encode a sort of *perspective* on a dimension *d*: they are exemplified only with respect to points of *d*, but they "tell" us something about what is going on in other points of the perspective. That is why there can be a trade off between the reality of *d*-tensed-properties and that of the (points of the) dimension *d*, and they are interesting with respect to the issue of the reality of the dimensions that *prima facie* we attribute to reality. We can take a *d*-tensed property to be nothing over and above a point of view on a dimension that is otherwise real *per se*, or we can take the dimension to be a sort of "illusion" induced by the instantiation of the *d*-tensed property, which is then seen as a genuine element of reality.[7]

Consider the case of *t*-tensed-properties, i.e. tensed properties. A so-called "tense realist" is a realist with respect to tensed properties.[8] Not surprisingly, tense realists are often anti-realist with respect to the temporal *dimension*, and they do not attribute reality to all temporal points in the same sense. For instance, the presentist maintains that only the present instant exist, while the growing block theorist maintains that only the instants up to the present exist. For the growing block theorist the temporal dimension does not spread into

[6] I am not the first to advance an analysis of tensed properties in terms of a simple "radical" and a tensed element (see Salmon, 2003; Crisp, 2007; Correia & Rosenkranz, 2011). I am sketching here a generalization from the case of time to those of other dimensions.

[7] Since any tensed property is constituted by a "P" element and a "tens" element, one might think that the latter are genuine elements of reality – the idea being that what is composed cannot be fundamental (thanks to Giorgio Lando for pointing this out to me). If we have a problem with that, we can either take "P" and "tens" to determine the tensed properties rather than being its constituents, or insist that the fundamental facts of which the dimension is only a reflection are facts about the instantiation of tensed properties (whether they are fundamental or not).

[8] See Fine (2005). Beware the possibility of terminological confusion: while "modal realist" is a realist with respect to the points of the modal dimension (possible world), a "tense realist" is a realist with respect to tensed properties (and possibly an anti-realist with respect to the temporal dimension, as we shall see shortly).

the future, and for the presentist it does not spread into the past either: we can only metaphorically talk of a "complete" temporal dimension. However, tensed properties can play the theoretical role that instants play depending on who considers the temporal dimension real. In particular, truths about what happens at non-existing locations of the dimension can be grounded in tensed properties exemplified at existing points of the dimension. And the grounding will be explanatorily felicitous precisely because of the realist stance towards tensed properties: they are the *part of reality* that does the grounding. For example, according to the "Lucretian" version of presentism, the ground of all what was and all what will be the case is the present instantiation of past- and future-tensed properties.[9]

To the contrary, tense anti-realists are usually realists with respect to the temporal dimension. Attributions of tensed properties relative to a certain instant *i* should not be taken at face value because what grounds tensed truths is the exemplification of simple properties relative to instants that are possibly distinct from *i* (alternatively, it is the atemporal exemplification of the relevant *t*-indexed-properties). Thus, the fact that this apple in front of me now possesses the property *having been green* is nothing over and above the fact that it possesses the simple property of *being green* at some other earlier instant.

What about *m*-tensed properties? They are quite often simply called modal properties: *being possibly red* is an example. Should we take talk of modal truths as grounded on exemplification of modal properties in the actual world, or rather as grounded on exemplification of simple properties along the whole modal dimension? The modal realist is likely to choose the second option. How to spell out exactly the relation between modal truths and the modal dimension depends on other details: in particular, it will depend on how seriously the modal realist takes modal properties. According to the more radical position, modal talk is merely a reflection of the fact that we are "perspectively" located in the modal dimension. But there are less radical positions as well. Modal properties can be thought of as real, but not fundamental, because they are

[9] See (Bigelow, 1996). Of course, if the tense realist does not think that there are truths concerning certain non-existing parts of a dimension, she can do away with grounding altogether. For instance, according to a certain conception of the open future, there are no (contingent) truths about the future precisely because there is no ground for them (and necessary truths about the future may be seen as grounded in atemporal logical necessities). See, for instance, (Markosian, forthcoming).

metaphysically or conceptually reducible to non-modal ones. I will say some more about these options for the modal realist in the next paragraph. Here I wish to stress that in so far as the ground for the attribution of modal properties will be what happens at other points of the modal dimension, the modal realist takes modality as a real dimension.

The actualist, too, can correlate modal truths in terms of what happens at different points of the modal dimension, but since all points of the modal dimension depend on the actual world, the grounding of modal truths will be on some aspect or things in the actual world. If the actualist is also a realist with respect to modal properties (not to be confused with the "modal realist" above), she can ground modal truths on exemplification of modal properties, taken as primitive and irreducible (Forbes, 1985). However, this is not the only option; she may endorse a more deflationary stance towards modal properties and consider them as having a conceptual nature (Plantinga, 1974). In any case, the actualist is bound to deny modality the status of a real dimension, because none of its points other than the actual world can be the ground of the true attribution of modal properties.

5. Counterparts and Modal Wholes

If the concept of modality as a dimension of reality is conceptually cogent, material things can exist in different possible worlds (although maybe not in all of them). And if modality is a real dimension, possible worlds will all be ontologically on a par. Thus, at different points of the modal dimension, we will find material beings; but what about the possibility of finding the *same* material being at different points of the modal dimension? Famously, Lewis has argued that there is a crucial difference here between the modal realist and the actualist. According to the actualist, possible worlds are to be found in the actual world, and their material constituents (if they have any) will be part of the actual world. Hence, it is not problematic to think of the existence of an entity *x* in different possible worlds in terms of those worlds sharing *x* as a part. There is something metaphorical in such a notion of "existence in a possible world", but it is precisely the aim of the whole actualist talk about the modal dimension and its points not to take them ontologically too seriously.

In contrast, modal realism takes them seriously, but this has problematic consequences for the idea that possible worlds overlap (i.e. share parts). Indeed, Lewis's modal realism is characterized by the thesis that possible

individuals are world-bound, i.e. they cannot entirely-exist in more than one world – and hence possible worlds are mereologically disjoint. Indeed, possible individuals are concrete individuals, and there is nothing intrinsic distinguishing them from actual ones. At each point of the modal dimension, namely in each possible world, we can find a *counterpart* of an entity existing in some other world (Lewis, 1968). The counterpart relation is a relation of similarity: my counterpart in a possible world *w* different from the actual world @ that I inhabit is the individual who is most similar to me in *w*.[10] If we assume a completely unrestricted principle of mereological composition, counterparts of each other will also compose a whole, and hence be parts of it. Lewis endorses unrestricted composition, and thereby acknowledges the existence of such "modal wholes". Does it follow that the same material entities can exist in more than one world – at least in the loose sense of "the same"? Not quite. According to Lewis, modal wholes are *not* possible individuals, and thus (in so far as being a possible individual is a necessary condition for being a material entity), they are not material entities (although all their parts are material). Thus, there is no material entity that is partially located at different points of the modal dimension, and it cannot be the case that the same material entity – in the loose sense (and of course in the strict sense) – exists at more than one possible world.

It is not inconsistent to drop the restriction on possible individuals (to the effect that modal wholes are not possible individuals) and to maintain that modal wholes are material entities. By endorsing such a position we would be go a step further in taking modality seriously as a dimension. Material entities extend through *m*, and have *m*-parts at different points of *m*, and thus we can find the same entity at different points of the modal dimension. More to the point, if modal wholes are material entities along the modal dimension, modal properties can be construed as perspectival effects, in complete analogy to what happens in the spatial dimension (and in the temporal dimension, according to the four-dimensionalist). Lewis indeed discusses the possibility of accepting "modal continuants".[11] However, he did not take this step, and here is his main argument against it:

[10] For simplicity, I am assuming that any entity has one counterpart in each possible world. On that see Lewis 1968.

[11] See also (Varzi, 2001), (Schlesinger, 1984) and (Begolo, ms). In (Simons, 1987, p. 361), the possibility of having a fifth modal dimension in that sense is hinted at but then dismissed.

> [...] even if Humphrey is a modal continuant, it doesn't take the whole of him to do such things as winning or thinking [that he might have won]. The continuant does them by having a world-stage that does them [...], just as a temporal continuant does this or that through its stages. But what good is that? If Humphrey yearns to think only of himself and nobody else, it is no use that he the thinker is part of the same mereological sum as some winner. That much is provided by the thiswordly sum of Humphrey and Nixon! No; what matters is that the modal continuant is counterpart-interrelated, so that the thinker of the thought has a winner for a counterpart. Then why not just say so, and leave the modal continuant out of it? (Lewis, 1983, p. 42)

I do not aim at an exegesis of Lewis's text here. He seems to claim that, assuming modal realism, in any account of modal properties the counterpart relation will be explanatorily more fundamental than the part-whole relation. If that is true, then the whole enterprise of reducing modal properties to "perspectival" effects on the modal dimension looks hopeless (which might not have worried Lewis). Modal realism can provide the framework for various kinds of "reductions" of modal properties. A metaphysical reduction is carried about by considering modal properties as "hidden" relational properties along the following lines: attributing a modal property P-*tens* is tantamount to attributing any of certain related *m*-indexed properties. For instance, an entity *x* is *possibly red* in the actual world if and only if *x* exemplifies any of the properties of the form *being-red-at-w* for some world *w*. This metaphysical reduction does not require counterpart theory. With counterpart theory we can go a step further and reduce *m*-indexed properties to simple properties. An entity *x* in a world *w'* has the P-*w* if and only if it has a counterpart in *w* that has P. *M*-indexed properties can thus be construed as nothing over and above simple properties possessed by counterparts in the appropriate places of the modal dimension.[12] But why should the counterpart relation be *relevant* to accounting for the attributions of modal properties that we make in the actual

[12] Can this metaphysical reduction also be construed as a conceptual reduction? Namely, is it plausible to think that each piece of reasoning that involves modal concepts could be carried out in an entirely non-modal conceptual framework? David Lewis, at least at certain moments, seems to think so: "Modal reasoning can be replaced by non-modal, ordinary reasoning about possible things" (Lewis, 1970, p. 175). Of course, one may suspect that "possible things" in the above quote is indeed the essentially modal ingredient that spoils the reductionist project. But if the counterpart relation can be further exploited to reduce such a notion to clearly non-modal ones, then the objection misfires.

world?[13] If what Lewis says against the theory of modal occurrents is correct, it is the *similarity* between counterparts that does the explanatory job.

Normally, we consider what happens to individuals similar to us as relevant to what *could happen* to us. Consider the role of statistics: if I know that a high percentage of people who smoke two packs of cigarettes every day, eat junk food and drink large amounts of alcohol will have heart attacks in their forties, and if I am in my forties and lead a similar lifestyle, I am justified in thinking that it is *possible* for me to have a heart attack, and indeed, I am justified in thinking that it is even likely. The similarity between the counterparts seems to play an extra role that the mere fact of being connected in a material whole cannot play. The idea, then, is that the modal continuant theory without counterpart theory just posits the existence of modal wholes as grounds of attributions of modal properties, while continuant theory *with* counterpart theory is just counterpart theory with an idle extra part. If we accept explanation of attribution of modal properties in terms of similarity to possible individuals, we do not need to accept modal continuants as well.

However, it is difficult to understand why similarity would be relevant for attributions of modal properties if modal properties did not depend on the actual simple properties of the object that we are considering. If possible objects are just concrete objects like the actual ones that we meet in ordinary life, and do not enjoy any peculiar status (in particular, no primitive modal property), the mere fact that there is a similarity relation between me and a possible individual cannot tell me anything about my modal status. In many cases, the fact that I stand in a similarity relation to something possessing certain (non-modal) properties will tell me something about my intrinsic constitution, and *if* that is relevant for what is possible or necessary for me, then it will also tell me something about my modal status – otherwise it will not. Hence, if we do not rely on some different project of reduction of modal properties to non-modal ones,[14] it is not clear what the advantage of similarity over the part-whole relation that "sticks" concrete modal wholes together could ever be. Yet, if the alternative to grounding modality in similarity is

[13] This is a version of the famous Humphrey objection.

[14] Think about reductionist projects of modality that are compatible with actualism. For instance: the Diodorean reductionist project, pursued to a certain extent by Prior, which aims to reduce modal notions to temporal ones, or the statistic reductionist project, which aims to reduce modality to statistical distribution.

grounding it in brute facts about modal wholes, then it is not clear whether modal continuants can provide a ground for modal properties at all.

This situation may complicate the whole idea of construing modality as a dimension of reality. However, even if the project of construing modality as a real dimension turns out to be hopeless, it still merits investigation in order to clarify the general constraints for being a dimension of reality along space and time. Even if modality does not, we may discover that something else passes the test.[15]

REFERENCES

Balashov, Y. (2010). *Persistence and Spacetime*. Oxford: Oxford University Press.

Begolo, N. (2010). *Fivedimensionalism*. University of Verona, Tesi di Laurea Specialistica in Filosofia (MA thesis in Philosophy).

Bigelow, J. (1996). Presentism and Properties. In J.E. Tomberlin (Ed.) *Philosophical Perspectives*, 10. Malden (MA): Blackwell, 35–52.

Calosi,C. (ms). Metaphysics of Persistence and Unrestricted Composition. Manuscript.

Casati, R., & Varzi, A. (1994). *Parts and Places: The Structure of Spatial Representation*. Cambridge, MA: The MIT Press.

Crisp, T. (2007). Presentism and Grounding. *Nous*, *41*, 90–109.

Correia, F., & Rosenkranz, S. (2011). Eternal Facts in an Ageing Universe. *Australasian Journal of Philosophy* (published online first, 30 June 2011, http://www.tandfonline.com/doi/abs/10.1080/00048402.2011.588952).

Fine, K. (2000). A Counter-example to Locke's Thesis. *The Monist*, *83*(3), 357–361.

[15] For instance, (Lockwood, 2005) has argued that the many-mind interpretation of quantum mechanics that he defends suggests the existence of a dimension of "actuality".

Fine, K. (2005). Tense and Reality. In K. Fine, *Modality and Tense. Philosophical Papers*. Oxford: Oxford University Press, 261–320.

French, S. (2011). Identity and Individuality in Quantum Theory. In Edward N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2011 Edition).

Forbes, G. (1985). *The Metaphysics of Modality*. Oxford: Oxford University Press.

Hawley, K. (2001). *How Things Persist*. Oxford: Oxford University Press.

Heller, M. (1990). *The Ontology of Physical Objects*. Cambridge, UK: Cambridge University Press.

Leonard, H.S., & Goodman, N. (1940). The Calculus of Individuals and Its Uses. *Journal of Symbolic Logic*, *5*, 45–55.

Lewis, D. (1968). Counterpart Theory and Quantified Modal Logic. *Journal of Philosophy*, *65*, 113–126.

Lewis, D. (1970). Anselm and Actuality. *Nous*, *4*, 175–188.

Lewis, D. (1983). Postscripts to "Counterpart Theory and Modal Logic". In D. Lewis, *Philosophical Papers, Vol. I*. Oxford: Oxford University Press, 39–46.

Lewis, D. (1986). *On the Plurality of Worlds*. Cambridge: Blackwell.

Lockwood, M. (2005). *The Labyrinth of Time*. Oxford: Oxford University Press.

Lowe , E. J. (1986). On a supposed temporal/modal parallel. *Analysis*, *46*(4), 195–197.

Markosian, N. (forthcoming). The Truth about the Past and the Future. In F. Correia, & A. Iacona (Eds.), *Around the Three*. Berlin: Synthese Library, Springer Verlag.

Oaklander, N.A. (2004). *The Ontology of Time*. Amherst, NY: Prometheus Book.

Over, D.E. (1986). On a Temporal Slippery Slope Paradox. *Analysis*, *46*(1), 15–18.

Plantinga, A. (1974). *The Nature of Necessity*. Oxford: Oxford University Press.

Rodriguez-Pereira, G. (2004). The Bundle Theory is compatible with distinct but indiscernible particulars. *Analysis*, *64*(1), 72–81.

Salmon, N. (2003). Tense and Intension. In A. Joickic, & Q. Smith (Eds.) *Time, Tense, and Reference*. Cambridge, MA: The MIT Press.

Schlesinger, G. (1985). Spatial, Temporal, and Cosmic Parts. *Southern Journal of Philosophy*, *23*(2), 255–271.

Sider, T. (2001). *Four-Dimensionalism. An Ontology of Persistence and Time*. Oxford: Oxford University Press.

Simons, P. (1987). *Parts. A Study in Ontology*. Oxford: Oxford University Press.

Varzi, A. (2001). Parts, Counterparts, and Modal Occurrents. *Travaux de logique, 14*, 151–171.

Williams, J.R.G. (2010). Fundamental and Derivative Truths. *Mind*, *119*(473), 103–141.

Commentary

David Lewis
Counterfactuals
Oxford, Blackwell, 1973 *

Vittorio Morato †
vittorio.morato@unipd.it

Counterfactuals, published in 1973, was the culmination of work (both technical and philosophical) that Lewis had done in previous years (see Lewis (1971), Lewis (1973b)). The bulk of his analysis is that counterfactuals are some sort of *variably* strict conditionals, whose semantics can be given in terms of (ordered sets of) possible worlds.

After the publication of *Counterfactuals*, the only articles Lewis wrote directly about the semantics of counterfactuals were Lewis (1977), Lewis (1979), and Lewis (1981). The first is a defense of Lewis's approach for counterfactuals with disjunctive antecedents (more on this below); the second is a full development (after an embryonic presentation in the book) of Lewis's approach to *backtracking counterfactuals*, conditionals of the form "if A were to happen at time T_A, then B would happen at time T_B"; the third is a comparison (indeed, a proof of equivalence) of Lewis's approach with that developed by A. Kratzer in 1981, where the factual background against which a counterfactual is evaluated is a set of premises, rather than an ordering of worlds. The principal changes to the 1986 "revised printing" edition consisted just of some corrections in the completeness results of chapter 6, plus other minor corrections. The core semantic analysis presented in *Counterfactuals* thus remained quite stable from its first publication onward.

While squarely a book about the semantic analysis of counterfactuals (at the time, more often called "subjunctive conditionals"), the book frequently steps outside this area to discuss topics in metaphysics, philosophy of language and philosophy of science: there are in fact many interesting passages about the

* Based on the revised printed edition, 1986.

† Department of Philosophy, University of Padua, Italy.

metaphysics of possible worlds, an analysis of the notion of law of nature and discussions about the vagueness, contexts and applications of the semantic approach developed for counterfactuals to other notions like conditional obligation, temporal notions and contextually definite descriptions. The style of the book, however, is – as it is typical of Lewis – very concise, crisp and forthright (it is less than 150 pages).

The book is organized into six chapters. The first chapter (nearly a third of the whole book) is a presentation of the semantics, while chapter 2 presents some "alternative reformulations". In chapter 3, Lewis compares his analysis with what he calls "metalinguistic theory" – a term he uses to cover the theories of Chisholm, Goodman and Mackie – and with Stalnaker's approach. Chapter 4 is a discussion of the two fundamental conceptual blocks of the theory, namely possible worlds and comparative similarity. In chapter 5, Lewis studies the presence of variably strict conditionals in other areas of intensional logic, namely deontic logic (conditional obligation), temporal logics (constructions like "when next" and "when last"), and "egocentric logic" (a term due A. Prior (1968) – quite unusual nowadays – that basically covers pre-Kaplanian attempts to develop a logic for indexicals). Finally, chapter 6 presents a regimentation of the materials into formal systems, for which Lewis proves completeness and decidability results. In the revised printing edition, the book ends with an appendix that contains a bibliography of related work, annotated by Lewis himself.

With respect to the philosophical background at the time of publication and abstracting from the technicalities that will be discussed in detail below, Lewis's approach is characterized by two theses:

- Counterfactuals have truth conditions;
- The truth-conditions of counterfactuals could be given in terms of possible worlds.

Surely, neither of these two features were taken for granted at the time of publication (nor, in some respects, are they even today);[1] especially the possible worlds analysis was quite novel. At that time, the landscape was still dominated by Goodman's "cotenability" approach. According to the cotenability approach, a counterfactual conditional of the form $\varphi \ []\rightarrow \psi$ is

[1] See (Barker, 2011) for a very recent attempt to criticize possible worlds semantics for counterfactuals.

true if and only if ψ can be derived (according to some laws) by φ and some other contingently true propositions. On this view, counterfactual conditionals are not really sentences – not entities to which truth conditions can be assigned – but rather elliptical presentations of arguments.

The possible worlds approach to counterfactuals, however, was not original with Lewis. W. Todd (1964) was probably the first author to lay the groundwork for such an analysis, as this quotation (p. 107) quite explicitly reveals:

> When we allow for the possibility of the antecedent's being true in the case of a counterfactual, we are hypothetically substituting a different world for the actual one. It has to be supposed that this hypothetical world is as much like the actual one as possible so that we will have grounds for saying that the consequent would be realized in such a world.

This line of research was then fully developed by Stalnaker (1968) and, more formally, by Stalnaker and Thomason (1970). As Stalnaker himself acknowledged, his approach and Lewis's, though quite similar, were developed independently from one another.[2]

The basic idea of Lewis's analysis is well expressed in the very first paragraph of the book:

> "If kangaroos had no tail, then they would topple over" seems to me to mean something like this: in any possible state of affairs in which kangaroos have no tails, and which resembles our actual state of affairs as much as kangaroos having no tails permits it to, the kangaroos topple over.

This paragraph contains *in nuce* almost all essential elements of Lewis's analysis. Let us start then from here to develop, step by step, a complete and more precise formulation of the truth-conditions for counterfactuals.

The general idea is that a counterfactual of the form $\varphi \ []\rightarrow \psi$ is true, in a possible world *i*, if and only if, in any world where φ is true and that resembles *i* as much as the truth of φ permits it to, ψ is true. We will see in a moment what it means that a world resembles another *as much as the truth of a proposition* φ *permits it to*. The important thing to notice, for now, is a structural feature. The form of this preliminary formulation of the truth-conditions for counterfactuals is the following:

[2] Stalnaker's acknowledgment is reported by Bennett (1974, p. 308).

φ []→ ψ is true in *i* if and only if for any world *w*, such that _____, ψ is true in *w*.

where "_____" stands for a restrictive condition that we will analyze in a moment. The left-hand side of these truth-conditions is structurally similar to the left-hand side of the truth-conditions of another, quite familiar, formula, namely □ψ:

□ψ is true in *i* if and only if, for any world *w* such that _____, ψ is true in *w*.

In the case of the truth-conditions for □ψ, the restrictive condition is filled up by the specification of an accessibility relation between *i* and the possible worlds universally quantified over.

In the case of the truth-conditions for φ []→ ψ, the restrictive condition is to be filled up by (the formal translation of) a condition having to do with *i* and φ. This condition is "being a world that resembles *i* as much as the truth of φ permits it to". It is then expected that such a condition determines a class of worlds. The counterfactual is then said to be true if and only if ψ is true in such a class of worlds.

Given that the condition that restricts the universal quantification in a counterfactual conditional is defined relative to the antecedent of the counterfactual (and to the world of evaluation, but this is true also for the accessibility relation used to restrict the quantification over worlds in the clause for □), a nice way, suggested by R. Stalnaker, to present Lewis's approach is to say that in his analysis the antecedents of conditionals "act like necessity operators on their consequents".[3] We will see how this analogy will reveal itself to be very useful for settling semantic issues for counterfactuals.

The class of worlds determined by φ and *i* is such that in it there are no worlds where φ is true and ψ is false. We could then say that a counterfactual like φ []→ ψ is true in *i* if and only if the material conditional φ → ψ is true in every member of the relevant class of worlds. We could slightly change our clause for []→ in order to register this new information:

φ []→ ψ is true in i if and only if, for every possible world *y* such that resembles *i* as much as the truth of φ permits it to, φ → ψ is true in *y*.

[3] For this view, see Stalnaker (1978, p. 93).

This clause, however, has two problems: the first is that it still contains the too informal and unexplained phrase "such that resembles *i* as much as the truth of φ permits it to"; the second is that the left-hand side of the biconditional could be vacuously satisfied in the case where there are no possible worlds that resemble *i* as much as the truth of φ permits them to, typically in the case where φ is impossible. The first problem could be solved by transforming the phrase "such that resembles *i* as much as the truth of φ permits it to" into the more tractable "such that φ is true in *y* and *y* resembles *i*", where *y* is bound by the universal quantifier. The new clause is surely more tractable, even if less expressive than the original: if φ is true in a world *x* that resembles *i*, then *x* is a world that resembles *i* as much as the truth of φ permits it to. Being true in a world is no conventional matter, so if something is true in a world, it is "permitted" to be true in such a world. For Lewis, the role of the informal clause (and in particular the use of "permits") was that of highlighting the fact that the relevant worlds to consider when evaluating in *i* a counterfactual like φ $\Box\!\!\rightarrow$ ψ are not those where φ is true and *everything else* is as it is in *i*. The reason is that, for Lewis, there are no such worlds. Or better, these worlds would be surprisingly far different from the actual, so different as to become irrelevant for the evaluation of the counterfactual. For example, the worlds where kangaroos have no tails and everything else is as it actually is are worlds less similar to the actual world than are the worlds where there are series of further deviations from actuality that "accommodate" the absence of tails in kangaroos in such worlds (due to a difference in the genetic set-up of kangaroos, for example). Here is what Lewis writes with respect to the similarity and difference trade-off:

> Respects of similarity and difference trade-off. If we try too hard for exact similarity to the actual world in one respect, we will get excessive differences in some other respect. (Lewis, 1973a, p. 9)

We should not expect, however, that, in translating an informal condition in quasi-formal terms, every aspect of the informal idea will be explicitly preserved. As far as the truth-conditions of counterfactuals are concerned, we can live with the bare-bones formulation given in terms of truth of φ and similarity to *i* and leave more sophisticated features to the informal interpretation of our primitives (in our case, the similarity relation between worlds). In light of this, we can now write again a new formulation of our truth-conditions:

> $\varphi \; []\!\rightarrow \psi$ is true in *i* if and only if, for every possible world *y* such that φ is true in *i* and *y* resembles *i*, ψ is true in *y*.

The second problem (i.e., the eventual vacuous truth of the left-hand side of the biconditional) is not a problem *per se*; a distinctive feature of Lewis's semantics for counterfactuals is that counterfactuals with impossible antecedents are true. Here is what Lewis writes:

> Confronted by an antecedent that it is not really an entertainable supposition one might react with a shrug: If that were so, anything you like would be true. (Lewis, 1973a, p. 23)

Furthermore, Lewis claims, counterfactual conditionals with impossible antecedent are asserted, by way of *reductio*, in philosophical, logical or mathematical arguments and need therefore to be taken as true in those contexts.

Note that the case of vacuous truth for $\varphi \; []\!\rightarrow \psi$ could again be seen in analogy with what happens in the case of $\Box\psi$. In a world *i* such that it is not in a relation of accessibility with any other world (nor with itself), every formula is necessary: given an arbitrary, $\Box\psi$ is true in *i*. Analogously, in a world *i* that is not in the relation of accessibility with any world where φ is true, for any ψ, every counterfactual of the form $\varphi \; []\!\rightarrow \psi$ is true.

What we want our truth-conditions to reveal, however, is what happens in the "normal" cases, those where φ is an entertainable supposition, namely where φ is possible. But inserting explicitly the possibility of φ in the truth-conditions for counterfactuals is going to add some complications. In particular, we would have to deal with three variables for worlds (the variable for the world of evaluation *i*, the variable introduced by the restricted universal quantification and the new variable, existentially quantified, for the world where φ is true), but with a relation of similarity with only two places.

The problem could be solved by introducing a three-place relation of similarity among worlds, *comparative similarity* (see Lewis, 1973a, p. 48):

> $x \leq_i y =_{\text{def}}$. *x* is at least as similar to *i* as the world *y* is.

Before discussing some properties of $\leq_i$, let us finally give the final formulation of the truth-conditions for counterfactuals in terms of this new relation of comparative similarity:

> $\varphi \; []\!\rightarrow \psi$ is true in i if and only if either

- *Vacuous case*: φ is false in any world accessible from *i* (i.e., φ is impossible); or
- *Non-vacuous case*: there is a world *x* accessible from *i* such that φ is true in *x* and, for any world *y*, if $y \leq_i x$, then ψ is true in *y*.

A counterfactual is (non-vacuously) true in a world *i* if and only if, if there is at least an accessible world *x* where the antecedent is true, then the consequent is true in every world at least as close to i as it is *x*.

In order for the truth-conditions to be working, we have to assume, as usual, that the standard binary relation of accessibility *R* and the new two-place relation $\leq_i$ are defined for any possible world *i* with respect to any possible world. The latter notion will then generate an ordering of all possible worlds with respect to their comparative similarity to *i*. In order to see what kind of ordering is generated by $\leq_i$, we need to know its properties:

- $\leq_i$ is *transitive*: whenever $x \leq_i y$ and $y \leq_i k$, then $x \leq_i k$;
- $\leq_i$ is *strongly connected*: for every *x* and *y*, either $x \leq_i y$ or $y \leq_i x$.

I assume that the role and meaning of transitivity is clear. The role of strong connectivity is that of assuring the possibility of comparisons (with respect to *i*) of any arbitrary pair of possible worlds. Given that strong connectivity entails reflexivity, this condition implies also that i is at least as close to itself than any other world is.[4]

To these conditions on $\leq_i$, at least these other two features should be added:

- Every possible world *i* is accessible to itself (i.e., *R* is a reflexive relation);
- Every possible world *i* is "strictly minimal" with respect to $\leq_i$, namely for any world *x* (different from *i*), $i <_i x$ (*i* is more similar to itself than any other world is).

The strict minimality condition is responsible, in Lewis's approach, for the fact that counterfactuals with true antecedents might be true. What happens is that counterfactuals with a true antecedent reduce to material conditionals (see (Lewis, 1973a, p. 26)). In order to see this, suppose that φ is true at *i*; then

[4] In passing, note that the relation of comparative similarity between worlds used to evaluate counterfactuals is slightly different from the counterpart relation of similarity used to evaluate de re modal claims: in particular, the counterpart relation is a non-transitive similarity relation and is also non-symmetric.

there is a world accessible from i (i itself, given that it is assumed that the relation of accessibility is reflexive) where φ is true. If ψ is true in i, then in every world as close to i as i is to itself, $\varphi \rightarrow \psi$ is true. This fact is granted by the strict minimality condition that assures that there is no other world but i that is as similar to i as i. Therefore, where φ and ψ are true in i, $\varphi \; []\rightarrow \psi$ is true in i.

Contrary to what appears to be the case, it is a very welcome feature of Lewis's approach that "counterfactuals" with true antecedents turn out true. The term "counterfactuals" is in effect quite misleading, and if used nonetheless, it should not be used in a sense that presupposes or implies the falsity of its antecedent. As T. Williamson (2008, p. 137) rightly claims, in some contexts, we can assert a sentence like "if Jones had taken arsenic, he would have shown just exactly those symptoms which he does in fact show" to be abductive evidence (by inference to the best explanation) for its antecedent, "Jones took arsenic".

But the eventual later empirical discovery that in effect Jones took arsenic does not make the previous assertion of the counterfactual inappropriate, rather it would constitute evidence for it. The counterfactual is true because we discover its antecedent to be actually true. Counterfactuals are "counter"-factuals not because they imply the falsity of the antecedent, but because their evaluation requires comparisons of alternative possible situations with the actual ones.

It is also interesting to note what properties $\leq_i$ does *not* have. In particular, $\leq_i$ is not anti-symmetric. That $\leq_i$ is not anti-symmetric means that from $x \leq_i y$ and $y \leq_i x$, it does not necessarily follow that $x = y$: there might be two distinct worlds, none of which is more similar to i than the other. The non-anti-symmetricity of the $\leq_i$ relation is one of the two main features that distinguishes Lewis's approach from Stalnaker's.[5] The effect of this choice is

[5] The other is the falsity, in Lewis, of the "limit assumption", namely the thesis that, when evaluating the counterfactual $\varphi \; []\rightarrow \psi$ in i, there is always a closest to i world where φ is true; the effect of such an assumption is that when we know that a formula φ is true in some world, we know also that it is true in some world that is the closest world to i. This feature is very useful when proving the validity or the invalidity of counterfactuals. The falsity of the limit assumption implies that, when we have infinitely many worlds, there is no the closest to i world, but only an infinite series of closest and closest worlds. Note, however, that if the worlds are finite in number, the limit assumption automatically holds, even in Lewis.

that in Lewis (and not in Stalnaker) the following two schemas ("conditional excluded middle" and "distribution") are not valid:

$$(\varphi \ []\!\rightarrow \psi) \vee (\varphi \ []\!\rightarrow \neg\psi)$$
$$(\varphi \ []\!\rightarrow \psi \vee \chi) \rightarrow ((\varphi \ []\!\rightarrow \psi) \ \vee (\varphi \ []\!\rightarrow \chi))$$

A countermodel for the first schema is given by two worlds equally similar to *i*, say *x* and *y*, such that φ is true in both of them and ψ is true in *x* but false in *y*. Given that a counterfactual, to be true, has to be true in every world similar to *i*, none of the disjuncts of our formula is true in our model. A countermodel for the second schema is given by letting φ, ψ and $\neg\chi$ be true in *x* and φ, χ and $\neg\psi$ be true in *y*.[6]

The conditional excluded middle is taken by Lewis to be a very "plausible"principle, especially because it explains why, in natural language, we do not usually distinguish between external and internal negation of a conditional. The sentence "it is not the case that if you had walked on the ice, it would have broken" seems to us perfectly equivalent to "if you had walked on the ice, it would not have broken". Given that the conditional excluded middle is equivalent to $\neg(\varphi \ []\!\rightarrow \psi) \rightarrow (\varphi \ []\!\rightarrow\neg\psi)$ and everyone agrees that the converse, namely $(\varphi \ []\!\rightarrow\neg\psi) \rightarrow \neg(\varphi \ []\!\rightarrow \psi)$, is independently plausible, we have $\neg(\varphi \ []\!\rightarrow \psi) \leftrightarrow (\varphi \ []\!\rightarrow\neg\psi)$. The problem, for Lewis, is that, given a plausible (for him) definition of the "might" counterfactual $\varphi \ \Diamond\!\!\rightarrow \psi$ as $\neg(\varphi \ []\!\rightarrow\neg\psi)$, we can prove, by conditional excluded middle (now in the form $\neg(\varphi \ []\!\rightarrow\neg\psi) \rightarrow (\varphi \ []\!\rightarrow\psi)$), that $\varphi \ \Diamond\!\!\rightarrow \psi$ entails $\varphi \ []\!\rightarrow \psi$; given that the other direction is obviously true, what we have in effect proved is the equivalence between $\Diamond\!\!\rightarrow$ and $[]\!\rightarrow$. This is, of course, quite unwelcome, and it is basically the reason why Lewis gives up on this formula.

Failure of the distribution principle is quite understandable if we remind ourselves again of the similarities and between $[]\!\rightarrow$ and $\Box$ in Lewis's approach. The analogue distribution principle for $\Box$,

$$\Box (\varphi\vee \psi)\rightarrow \Box\varphi \vee \Box\psi$$

quite clearly fails (take ψ to be $\neg\varphi$, for example). Contrary informal intuitions for the validity of the distribution principle for counterfactuals could

[6] A nice formalism for establishing invalidity in counterfactual logic is presented in Sider (2010, p. 208–216).

then be taken simply as symptoms of "scope muddles" typical of interactions between disjunction and intensional operators. Similar "muddles" are probably in play also in the case of positive informal judgments on the validity of the following inferences, called "Simplification of disjunctive antecedent":

$$\frac{\varphi \vee \chi \; []\!\rightarrow \psi}{\varphi \; []\!\rightarrow \psi} \qquad \text{(DISJ-1)}$$

Or

$$\frac{\varphi \vee \chi \; []\!\rightarrow \psi}{\varphi \; []\!\rightarrow \chi} \qquad \text{(DISJ-2)}$$

The failure of Lewis's (and Stalnaker's) semantics to make such inferences valid is often presented as a drawback for both approaches and, in general, for any possible worlds approach to counterfactuals[1] (see (Nute, 1976); (Ellis, Jackson & Pargitter , 1977)). A countermodel to DISJ-1 in Lewis's semantics is done by supposing the existence of a χ and $\neg\varphi$ world w such that, in every x, such that $x \leq_i w$, is true in x; a countermodel to DISJ-2 by supposing the existence of a φ and $\neg\chi$ world, such that in every x such that $x \leq_i w$, ψ is true in x.

The problem is that we seem to be normally disposed to infer from "if φ or χ would have been the case, then ψ would have been the case" the conclusion that "if χ would have been the case, then ψ would have been the case" or "if φ would have been the case, then ψ would have been the case". For example, after a very boring evening spent at home watching TV, we could say something like "if we had gone to a cinema or to a theatre, it would have been definitely better", and from this it is quite natural to conclude also that "if we had gone to a cinema, it would have been definitely better" or "if we had gone to a theatre, it would have been definitely better". Accepting such schemas into the logic, however, would be particularly dramatic for Lewis. As pointed out in Fine (1975, p. 453), given the logical equivalence between φ and $(\varphi \wedge \chi) \vee (\varphi \wedge \neg\chi)$, from $\varphi \; []\!\rightarrow \psi$ we can derive by substitution of logical equivalents

$(\varphi \wedge \chi) \vee (\varphi \wedge \neg\chi)$ $[]\!\!\rightarrow \psi$, and then, by an application of DISJ-1, we can conclude $(\varphi \wedge \chi)$ $[]\!\!\rightarrow \psi$. But failure of this rule:

$$\frac{\varphi \; []\!\!\rightarrow \psi}{(\varphi \wedge \chi) \; []\!\!\rightarrow \psi} \qquad \text{(STRENGTH)}$$

is taken as a benchmark of Lewis's logic of counterfactuals. Failure of such a rule is one of the main arguments Lewis uses to prove that counterfactual conditionals are not to be understood as (and then not reduced to) special kinds of (non-variably) strict implications. Failure of STRENGTH is then essential for the identity of $[]\!\!\rightarrow$ as an autonomous type of conditional.[7]

One way in which the problem could be solved is by explaining away the evidence in favor of DISJ. In particular, what might be contested is the legitimacy of translating natural language counterfactuals with apparently disjunctive antecedents with a formula like $\varphi \vee \chi$ $[]\!\!\rightarrow \psi$. Interactions of other intensional operators with disjunctions in natural language might constitute a useful analogy. Take the case of permission. In natural language, from "it is permissible that φ or ψ" we seem quite naturally to be disposed to infer "it is permissible that φ and it is permissible that ψ". But adding the following rule to deontic logic:

$$\frac{P(\varphi \vee \psi)}{P\varphi \wedge P\psi} \qquad \text{(P-DISJ)}$$

would have dramatic consequences, because it would imply the truth of $P\varphi \rightarrow P\psi$; given the arbitrariness of φ and ψ, this formula would entail that, from the permission of doing something, anything is permitted (this problem is called "paradox of free choice permission"). One of the standard responses to such a case is to translate a sentence schema like "it is permissible that φ or ψ" not as its surface form would suggest, but rather like $P\varphi \wedge P\psi$. Similarly, one can adopt the same strategy for counterfactuals by translating natural language

[7] Actually, and more dramatically, if one agrees to contextually define $\Box\varphi$ as $\neg\varphi \; []\!\!\rightarrow\varphi$, then one can prove, by the essential use of DISJ and substitution of logical equivalents, that $\varphi \; []\!\!\rightarrow \psi$ entail $\varphi \rightarrow \psi$; see Loewer (1976, p. 532).

counterfactuals like "if φ or χ would have been the case, then ψ would have been the case" by φ []→ ψ ∧ χ []→ ψ. The reason for doing so would be mainly pragmatic: we seem to utter counterfactuals with disjunctive antecedents *only* in case we are *already* willing to assert the two simplified counterfactuals; we naturally tend to infer the two simplified counterfactuals from the disjunctive one only because it would be inappropriate to assert the latter without already believing the former. But then, the relation between the two simplified counterfactuals and the counterfactual with a disjunctive antecendent is not one of entailment, but of presupposition, and therefore there is no reason to consider DISJ, in its original formulation, a logically valid rule of inference.[8]

The truth conditions given here in terms of a relation of comparative similarity are not those preferred by Lewis as given in the first chapter of *Counterfactuals*. In the book, the "official" truth-conditions are given in terms of a nested system of "spheres", where a sphere is a set of possible worlds within a certain degree of similarity to a given world. In this formulation, a counterfactual []→ is true in *i* if either no sphere around *i* contains a φ-world or some sphere contains a φ-world and no φ and ¬ψ world.

The two formulations are, however, perfectly equivalent (Lewis himself proves this; see Lewis, 1973a, p. 49–50), but I think that the one in terms of comparative similarity is to be preferred. It is surely more fundamental. This is because spheres must be ultimately explained in terms of comparative similarity. A sphere is a set of worlds whose members are more (or less) as similar to a given world than the non-members. Any condition on spheres – their being nested, centered around a world, closed under unions, etc.[9] – is thus justified on the basis of corresponding conditions on the comparative similarity relation.

Furthermore, as Fine (1975, p. 457) pointed out, the formulation in terms of $\leq_i$ has the advantage of being first-order, while the formulation in terms of spheres presupposes an assignment to each world of a set of sets of worlds. Use of $\leq_i$ would also allow the substitution of the intensional language of modal and counterfactuals operators with an extensional language containing just first-order quantifications over worlds. In the same way in which modal languages

[8] On this see (Bennett, 2003, p. 168–171).
[9] Cf. Lewis (1973a, p. 14).

containing $\Box$ and $\Diamond$ could be substituted by first-order non-modal languages containing quantifiers over worlds and a group of new interpreted predicates (see Lewis, 1968), counterfactual languages containing $[]\rightarrow$ and $\Diamond\rightarrow$ could be substituted by first-order non-modal languages containing the interpreted predicates R (for "... is accessible from ... ") and $\leq_i$ (for "... is as similar to ... as ...").

There is, however, at least an expository advantage that spheres have over comparative similarity. By using spheres, the difference between counterfactual conditionals and strict conditionals, as I said, a benchmark of Lewis's analysis, could very vividly (almost graphically) be made.[10] To a (constantly) strict conditional there is an assignment to each i of a single sphere of accessibility, while to a variably strict conditional (i.e., a counterfactual) there is an assignment to each i of a system of spheres. A counterfactual is non-vacuously true in case the corresponding material conditional is true in every world belonging to at least one sphere around i; on the contrary, the strict conditional is true if the corresponding material conditional is true in every world accessible-to-i. To the counterfactuals is assigned a structured space of metaphysical accessibility, whose structure is given by the similarity relation dividing this space into spheres; the space of accessibility of strict conditionals is instead simply left unstructured. The counterfactual "if the bush had not been there, the rock would have ended in the lake" is true in the actual world, even if there is an accessible world where the bush is not there and the rock does not end in the lake; what is needed for the counterfactual to be true is rather a special sub-group of accessible-to-i worlds such that in every world where the bush is not there, the rock ends in the lake.

Of course, we can tweak accessibility relations to detect just those worlds where the corresponding material conditional is true. Restriction of the accessibility relation used to evaluate strict conditionals could be obtained in various ways. One way would be to make the accessibility relation contextually determined. This strategy is at the heart of quite recent attempts to analyze counterfactuals as kinds of highly context-sensitive (non-variably) strict conditionals. According to Fintel (2001), for example, in the case of counterfactuals, the determination of the accessibility relation useful to

[10] See Lewis (1973a, p. 11).

evaluate the conditional, what he calls "the modal horizon", is contextually determined and dynamically changes throughout a discourse. Relativization of the accessibility relation to dynamically determined modal horizons could be quite useful to explain certain linguistic phenomena involving counterfactuals, like sensitivity to order of utterance. If in a conversation I first utter the sentence "If I had struck the match, it would have lit" and then "If I had struck the match in outer space, it would have lit", the first sentence seems to be true and the second false. But if, in another conversation, the order of utterance is reversed, the latter sentence still seems false, but now the former also seems false.[11] This is because the relevant modal horizon for the conversation is that of the first uttered counterfactual and the horizon of the second counterfactual is wider than the one of the first counterfactual. Being wider, it contains worlds (which remains accessible throughout a discourse) where I strike the match, but it does not light and where therefore the material conditional is false.

It is difficult to say whether these contemporary and sophisticated contextual approaches represent a real novelty or are simply notational variants of the "classic" approaches given by Lewis and Stalnaker. The author of this commentary suspects that they are. Already in the seventies, however, just envisaging the (at the time open) possibility of treating counterfactuals as contextually strict conditionals, Lewis defined such strategies "defeatist" because, he wrote: «It consigns to the wastebasket of contextually resolved vagueness something much more amenable to systematic analysis than most of the rest of the mess in that wastebasket» (Lewis, 1973a, p. 13). This quotation reveals quite nicely and from a quite different angle, in what consists Lewis's often celebrated "systematic philosophy"; Lewis's systematicity consists not only in his capacity to clarify a great number of problems belonging to different philosophical areas, but also, and more importantly, in his capacity to tackle such problems by means of robust theorizing.

[11] This is an example taken from Sider (2010, p. 225).

REFERENCES

Barker, S. (2011). Can Counterfactuals really be about Possible Worlds? *Nous*, *45*(3), 557–576.

Bennett, J. (1974). Counterfactuals and Possible Worlds. *Canadian Journal of Philosophy*, *4*(2), 381–402.

Bennett, J. (2003). *A Philosophical Guide to Conditionals*. Oxford: Oxford University Press.

Ellis, B., Jackson, F., & Pargitter, R. J. (1977). An Objection to Possible Worlds Semantics for Counterfactuals Logic. *Journal of Philosophical Logic*, *6*, 35–57.

Fine, K. (1975). Critical Notice of Counterfactuals. *Mind*, *84* (335), 451–458.

Fintel, K. von. (2001). Counterfactuals in a Dynamic Context. In M. Kenstowicz (Ed.), *Ken Hale: a Life in Language*. Cambridge, MA: MIT Press, 123–152.

Kratzer, A. (1981). Partition and Revision: the Semantics of Counterfactuals. *Journal of Philosophical Logic*, *10*, 201–216.

Lewis, D. K. (1968). Counterpart Theory and Quantified Modal Logic. *Journal of Philosophy*, *LXV*(5), 113–126.

Lewis, D. K. (1971). Completeness and decidability of three logics of counterfactual conditionals. *Theoria*, *37*, 74–85.

Lewis, D. K. (1973a). *Counterfactuals*. Oxford: Blackwell.

Lewis, D. K. (1973b). Counterfactuals and Comparative Possibility. *Journal of Philosophical Logic*, *2*, 418–446.

Lewis, D. K. (1977). Possible-Worlds Semantics for Counterfactuals: a Rejoinder. *Journal of Philosophical Logic*, *6*, 359–363.

Lewis, D. K. (1979). Counterfactual Dependence and Time's Arrow. *Nous*, *13*(4), 455–476.

Lewis, D. K. (1981). Ordering Semantics and Premise Semantics for counterfactuals. *Journal of Philosophical Logic*, *10*, 217–234. (Reprinted as ch. 6 of Lewis (1998))

Lewis, D. K. (1998). *Papers in Philosophical Logic*. Cambridge: Cambridge University Press.

Loewer, B. (1976). Counterfactuals with Disjunctive Antecendents. *Journal of Philosophy*, *73* (16), 531–536.

Nute, D. (1976). David Lewis and the Analysis of Counterfactuals. *Nous*, *10* (3), 355–361.

Prior, A. N. (1968). Egocentric Logic. *Nous*, *2*, 191–207. (Reprinted as ch. XVII of Prior (2003))

Prior, A. N. (2003). *Papers on Time and Tense* (new edition). Oxford: Oxford University Press.

Sider, T. (2010). *Logic for Philosophy*. Oxford: Oxford University Press.

Stalnaker, R. (1968). A Theory of Conditionals. In N. Rescher (Ed.), *Studies in Logical Theory*. Oxford: Blackwell, 98–112.

Stalnaker, R. (1978). A Defense of Conditional Excluded Middle. In W. Harper, R. Stalnaker, & G. Pearce (Eds.), Ifs (pp. 87–103). Dordrecht: Reidel.

Stalnaker, R., & Thomason, R. (1970). A Semantic Analysis of Conditional Logic. *Theoria*, 36 , 23–42.

Todd, W. (1964). Counterfactual Conditionals and the Presuppositions of Induction. *Philosophy of Science*, *31*(2), 101–110.

Williamson, T. (2008). *The Philosophy of Philosophy*. Oxford: Blackwell.

Commentary

David Lewis
On the Plurality of Worlds
Oxford: Blackwell, 1986 *

Louis deRosset †
louis.derosset@uvm.edu

David Lewis is one of the most influential philosophers of our age, and *On the Plurality of Worlds* is his magnum opus. *OPW*[1] offers an extended development and defense of the hypothesis that there are many universes, things of the same kind as the universe in which we all live, move, and have our being. Lewis calls these universes "worlds", deliberately recalling the notion of a "possible world" familiar from modal logic and the metaphysics of modality.

The title invokes the thesis of the book: there are pluralities of worlds, things of the same kind as the world we inhabit, differing only with respect to what goes on in them. Lewis sought in earlier work (Lewis, 1973, pp. 84–86) to offer a direct argument from common sense modal commitments to the existence of a plurality of worlds.[2] *OPW* offers a less direct argument. Here, Lewis supports the hypothesis by arguing that, if we accept it, we have the material to offer a wide range of analyses of hitherto puzzling and problematic notions. We thereby effect a theoretical unification and simplification: with a small stock of primitives, we can analyze a number of important philosophical notions with a broad range of applications. But the analyses Lewis proposes are adequate only if we accept the thesis that there are a plurality of worlds. Lewis claims that this is a reason to accept the thesis. In his words, «the hypothesis is serviceable, and that is a reason to think that it is true» (p. 3).

* Thanks are due to Roberto Ciuni for comments on an earlier draft, and for Terence Cuneo and Mark Moyer for discussion.
† Department of Philosophy, University of Vermont, USA.
[1] In what follows, I will abbreviate the title to *OPW*; unless otherwise indicated, page and section references are to this book.
[2] See Stalnaker, 1976 for criticism.

OPW contains four large chapters. The first chapter fleshes out the thesis that there exists a plurality of worlds, and offers Lewis's analyses of philosophically important notions in terms of worlds and their denizens. The second chapter articulates and responds to objections to the hypothesis and its accompanying analysis of necessity and possibility. The third chapter surveys and offers objections to competing conceptions of the nature of possible worlds. The fourth chapter is dedicated to exploring topics touching on the phenomenon of *de re* modality, and, in particular, defending Lewis's distinctive, counterpart-theoretic approach to that phenomenon.

The influence of *OPW* consists mainly in the adoption of Lewis's methodology, rather than his doctrines.[3] The thesis of a plurality of worlds is no exception. In the words of Ted Sider, a prominent contemporary sympathizer, «almost no one other than Lewis accepts it in its entirety» (Sider, 2003, p. 193). I will offer, then, only a brief characterization of the thesis of a plurality of worlds and the concomitant analysis of modality, and trace the course of some (but only some) of the objections to Lewis's defense that have been explored in the literature. I do not aspire to completeness. Instead, I will choose objections in an attempt to highlight important aspects of Lewis's methodology.

1. The Thesis of a Plurality of Worlds

According to Lewis's theory, there are many worlds, each one a thing of the same kind as our world. I will call things of this kind "cosmoses," to emphasize Lewis's distinctive account of their nature. Our cosmos is familiar: it comprises an entire spacetime and all its contents. This is the only cosmos most of us believe in. Lewis holds, however, that there are other, less familiar cosmoses, which similarly comprise entire spacetimes and all of their contents.[4] Our cosmos, the actual world, is just one cosmos among many, coexisting with the others.

[3] Interestingly, Lewis's counterpart theory is more widely accepted. See Merricks, 2003 for references and criticism.

[4] See pp. 1–2 and 69–81. Lewis's official view is that cosmoses are maximal mereological sums of entities that are related either spatiotemporally or by a relation suitably analogous to a spatiotemporal relation. Thus, other cosmoses comprise either entire spacetimes or entities analogous to spacetimes; see pp. 75–76 for this wrinkle.

There are also many individuals in this array of cosmoses. Some of those individuals are in our cosmos; most are not. The things of one cosmos stand in no spatial or temporal relations to the things of another, nor is there causal interaction between cosmoses; different cosmoses are causally and spatiotemporally isolated from one another. Indeed, Lewis thinks that the cosmoses are individuated by their spatiotemporal isolation. If a thing in a cosmos *x* bears spatial or temporal relations to a thing in a cosmos *y*, then *x* and *y* are on that account the same cosmos.

No individual is in more than one cosmos on Lewis's view. This point comports with his conception of a world as a spatiotemporally isolated cosmos. Perhaps ordinary individuals like Obama can be scattered, having non-contiguous parts that are at some spatiotemporal distance from one another. But plausibly they cannot have parts that bear no spatiotemporal relations whatsoever to one another. So an ordinary individual like Obama cannot be in two entirely separate, disjoint cosmoses.[5]

Lewis introduces a new class of relations that may obtain between individuals of different cosmoses. These *counterpart relations* are founded in relations of similarity among these individuals. Roughly, *x* is a counterpart of *y* iff *x* resembles *y* in relevant respects to a sufficient degree and no other individual in *x*'s cosmos resembles *y* more closely in relevant respects than *x*. Different counterpart relations are yielded by different specifications of which respects are relevant. For instance, if the only respect of resemblance that is relevant in a given context is biological species, then in that context the counterpart relation obtains between you and any human being in any cosmos.[6]

A more finely grained treatment of Lewis's discussion of counterpart relations would be extremely complicated. The important point for present purposes is that the class of counterpart relations is delineated solely by

[5] McDaniel (2004, 2006) describes a view that attempts to reconcile the idea that worlds are spatiotemporally isolated cosmoses with the idea that a single individual inhabits more than one cosmos. The view accepts that Obama does not have parts in disjoint cosmoses, claiming instead that Obama is wholly located in disjoint cosmoses.

[6] This explanation of the counterpart relation follows the treatment in (Lewis, 1968), except for the proviso "in relevant respects," which is added to preserve some of Lewis's claims about the admissibility of counterpart relations in *OPW* (pp. 88–89, 248–255); interestingly, Lewis neither repeats this explanation of a counterpart relation nor offers any very precise alternative characterization in *OPW*.

reference to similarity in non-modal respects, without recourse to modal notions. Closeness of resemblance and sufficiency of degree of resemblance are apparently non-modal notions. Thus, counterpart relations are ultimately explained in terms of similarity in non-modal respects, without reference to necessity or possibility. They provide suitable raw material, then, for a proposal to reduce modal claims to a non-modal basis.

2. The Analysis of Modality

Interest ontology. Why should we believe in it? Lewis argues in *OPW* that we should believe in it because, if we do, we can use it to offer analyses of modality, counterfactuals, properties, and the contents of psychological attitudes.[7] I will concentrate on the analysis of modality, since that is the focus of much of the book.[8] Lewis attempts to reduce modality by pairing modal claims with proposed analyses in non-modal terms. Lewis's theory separates modal claims into two kinds. The truth of the first kind, the *de re* modal claims, turns on the possibilities for particular individuals. The truth, for instance, of 'Bush might have lost the 2000 electoral vote' turns on what's possible for Bush. Lewis relies on counterparts to analyze *de re* claims. His analysis of the claim at hand is: there is a counterpart of Bush, an inhabitant of some cosmos or other, who lost.[9] This claim is no more modal than the claim that there is an opponent of Bush, an inhabitant of some state or other, who lost the vote. Modal claims of the second kind, the *de dicto* modal claims, do not turn on the possibilities for any particular individual. The truth, for example, of 'there might have been purple penguins' does not turn on how any particular individual might have been, but rather on the possible truth of a general claim to the effect that there are some purple penguins. Lewis's analysis of *de dicto* claims does not appeal to counterparts. In the case at hand, Lewis's analysis is:

[7] See §§1.2–1.5.

[8] Lewis's proposed analyses of counterfactuals, properties, and the contents of attitudes each take up a subsection of the first chapter. The rest of the book is concerned with the proper treatment of modality.

[9] See pp. 12–13. Corresponding to each pairing of a claim of the form 'it is possible that φ' with a non-modal analysis is an account of what's required for φ to be true at a cosmos. The truth of claims concerning particular individuals at a given cosmos is given by the properties of those individuals' counterparts (if there are any) in that cosmos. In this case, 'Bush lost the 2000 electoral vote' is true at a cosmos *w* iff Bush has a counterpart in *w* who lost a counterpart in *w* of the 2000 electoral vote.

there is a cosmos which includes some purple penguins.[10] This claim seems no more modal than the claim that there is a state whose inhabitants include some wealthy surfers.

So far, we only have pairings of modal claims with non-modal analyses for two particular sentences. It would be useful to have a general recipe for pairing a claim made in the modal idioms of natural language with a proposed non-modal analysis in terms of cosmoses and counterparts. Lewis provided such a scheme in Lewis, 1968. It often goes unremarked, however, that this scheme is repudiated in *OPW* (pp. 12–13).[11] Lewis there argues that the analysis of modal claims in terms of cosmoses and counterparts will have to be done on a case by case basis.[12] (He would still endorse the analyses offered for the two cases we have encountered.)

Many philosophers would think of this lack of systematicity as an objectionable feature of his theory. I am inclined, however, to cut Lewis some slack here. If Lewis gives us reason to be optimistic that, for every particular modal fact, there is an acceptable analysis of that fact in terms of cosmoses and counterparts, then he will have shown that all of modal reality can be described in non-modal terms. Providing a general recipe for giving such an analysis would be an impressive reason for optimism on this score, but we shouldn't assume at the outset that it is the only such reason.

3. Objection: The View is Ontologically Extravagant

Lewis's thesis of a plurality of worlds is ontologically extravagant when paired with his attempted reductions of modal claims. According to Lewis's view, since it is possible that there be purple penguins, there are purple penguins, albeit in other cosmoses. This is certainly a claim we ordinarily would deny.

[10] See pp. 12–13. Again, Lewis gives us an account of what's required for a generalization like 'there are purple penguins' to be true at a cosmos; the generalization is true at a cosmos just in case the cosmos contains some purple penguins.

[11] Fara and Williamson, 2005 (pp. 26–28) are an exception.

[12] Thus Lewis anticipates the conclusion reached by the objection, laid out in Divers, 1999 (p. 227), that Lewis's view can offer no general recipe for interpreting modal claims; Divers lays out the objection as a prelude to attempting a rebuttal, and so does not endorse the conclusion. Lewis's argument (pp. 11–13) turns on the difficulty of offering a systematic account of Humphrey's satisfying both '*x* is necessarily human' and '*x* is possibly nonexistent' without also satisfying '*x* is possibly both nonexistent and human'.

This ontological extravagance is often presented as a conflict between Lewis's theory and common sense. But it is no decisive objection to a theory that it conflicts with common sense. The progress of science, especially physics, has shown that common sense may sometimes be set aside. Scientists have discovered, for instance, that, contrary to common sense, simultaneity is frame-relative: there is no such thing as two events that are related by absolute simultaneity. But the ontological extravagance of Lewis's theory puts it into conflict with more than just the deliverances of common sense. Given that it is possible that classical Newtonian physics have been true, Lewis's theory commits us to the claim that there is such a thing as two events that are related by absolute simultaneity, albeit in another cosmos. Lewis's view, then, appears to conflict not just with pre-theoretic common sense, but also with the deliverances of mature science. The apparent conflicts are best understood as conflicts between Lewis's view and common sense *aided and corrected by mature science*. The examples of such ontological extravagance can be multiplied.

Lewis was well aware of these conflicts, and acknowledged that this was an objectionable feature of his theory of possible worlds.[13] He claimed, however, that the other virtues of his view made it the most attractive alternative on balance. In particular, the fact that it facilitates the reduction of modality to non-modal terms was for Lewis a conspicuous virtue. Lewis ultimately concedes the objection from ontological extravagance. He characterizes the issue between his view and competing alternatives as a dispute about the appropriate way to balance the theoretical cost imposed by the view's ontological extravagance against the virtue of its reduction in the number of primitive notions.[14] Lewis holds that analyzing modality is worth the ontological extravagance of his view; reduction is more important than even this very severe form of ontological extravagance. His opponents disagree. One critic terms the ontological extravagance of Lewis's view «an appalling

[13] See p. 135. Lewis also attempts here to mitigate the ontological extravagance by arguing that the physicist's denial of the existence, e.g. of absolutely simultaneous events is consistent (when interpreted correctly in context) with his view. His response, in essence, is that when physicists say «No two events are absolutely simultaneous», they ordinarily mean that no two *actual* events are absolutely simultaneous. On Lewis's semantics for "actual" (§1.9, pp. 92–96), a proponent of his view can comfortably deny that there actually are any absolutely simultaneous events, *i.e.*, that there are any in our cosmos. See also the response to this problem at Lewis, 1973 (pp. 86–87).
[14] See p. 156.

violation».[15] Perhaps this is one reason why few contemporary Lewis enthusiasts endorse the thesis of a plurality of worlds: they think the costs in ontological extravagance outweigh the benefits of eliminating primitive modality. Nevertheless, the methodological lesson here is clear and has been influential: *Reducing the number of primitives matters a lot.* Common sense counts when assessing a theory, especially when aided and corrected by mature science, but reducing the number of primitive notions often counts for more.

4. Objection: The View is Unmotivated

Traditionally, reductive theories of modality have been motivated by epistemological or metaphysical concerns. One rough-and-ready epistemological motivation for reducing modal facts to non-modal facts is that our standard techniques for gathering evidence about, e.g., Obama only provide information about how he is; our observational and perceptual evidence gives us no information about how he might have been (other than the trivial information we glean by observing how he is). Similarly, our observational and perceptual evidence gives us no information concerning unactualized *de dicto* possibilities, like the possibility that there be purple penguins.[16] How, then, can we know about unactualized possibilities? If we can reduce unactualized possibilities to non-modal matters of fact, then we can know about the former in the same way we know about the latter.

Insofar as one finds these epistemological worries compelling, one will find Lewis's view objectionable. Lewis-style reductions just reintroduce the epistemological problem. Unactualized possibilities for Lewis turn on facts concerning how matters are in cosmoses which we do not observe or perceive. Lewis's response in effect is to reject the claim that our standard techniques for gathering evidence about Obama or penguins exhaust our techniques for gathering such evidence.[17] In particular, we do not need observational or perceptual evidence to give us information about things that exist necessarily, including numbers, sets, and possibilia. I won't pause to assess the plausibility of Lewis's response. The important point for present purposes is that Lewis clearly does not think that the desirability of reducing modality depends on the

[15] Melia, 2008, p. 136.
[16] See the historically important argument at Hume, 1739/2001 (§1.3.14.1).
[17] See Lewis, 1973 (p. 87) and *OPW* (§2.4, pp. 108–115).

idea that a reduction would secure the observability or perceivability of modal facts. His reductions are not epistemologically motivated.

Sometimes philosophers take reduction to serve metaphysical rather than epistemological ends. These thinkers motivate reductive programs as a means of solving metaphysical puzzles concerning the facts or notions to be reduced. Sider, for instance, argues that modal properties are suspect because they «point beyond themselves»: their instantiation involves more than «what objects are actually like». Sider calls properties that are suspect because they «point beyond themselves» *hypothetical properties*. He terms properties that do not «point beyond themselves» *categorical properties*.[18] A reduction of modal properties, according to Sider, would show how they are instantiated in virtue of some congeries of categorical properties, thereby removing the putative grounds for suspicion. Lewis, by contrast, offers no such reason. His reductions are not metaphysically motivated.

In fact, they are not motivated at all. The methodological principle here seems to be that *reduction needs no motivation*. Lewis's official view is that reduction is desirable even when it isn't motivated by any feature in particular of the reduced claims or facts. Fewer primitives make for a better theory, even if the reductions in question solve no particular epistemological or metaphysical problems. Solving such problems is at best a further factor counting in favor of a reductive proposal.[19]

5. The Relevance Objection

A further problem for Lewis's view is that goings-on in places spatiotemporally isolated from this cosmos appear to lack the right sort of relevance to the question of what might have been the case here. For instance, the loss of some electoral vote by someone else in a different cosmos, bearing no spatiotemporal relations to Bush at all, seems irrelevant to whether Bush might have lost the electoral vote here in 2000. It is implausible to hold that the victories of other individuals in other elections in other spacetimes is intimately

[18] See Sider, 2001 (p. 41) and Sider, 2003 (pp. 184–185).

[19] In this connection, it is worth emphasizing that Lewis (1973, p. 87) claimed that parsimony in one's stock of primitive notions counts for a lot, but parsimony in one's stock of entities counts for little or nothing. For instance, Lewis thinks it does not tell against his theory that it commits him to the existence of more human beings than rival theories that do not endorse the thesis of a plurality of worlds. See Nolan, 1997 for criticism on this point.

linked with the fact that Bush might have lost here in the way required by Lewis's analysis. The point is reinforced by a comparison to elections in other places in the cosmos which we all inhabit. The losses of other individuals in other elections in other countries seem to have nothing to do, in the relevant sense, with the possibility of a Bush loss here.

It's an overstatement to suggest that the losses by other people in other elections have *nothing at all* to do with the possibility of a Bush loss here. The loss of someone else elsewhere might provide *evidence* that Bush might have lost. But this is not the sort of relevance at issue. According to Lewis's view, part of what it is for Bush to be a possible loser is that someone elsewhere who resembles Bush is a loser in fact. The objection appeals to what we might call a *constitutional intuition*: an intuition concerning what the modal fact in question consists in. According to the objection, Bush's being a possible loser does not consist, even in part, of someone else's losing. Since Lewis's view says otherwise, the objector argues, Lewis's view is incorrect. The evidential relevance of someone else's losing an election somewhere else does not impugn the cogency of the objection.[20]

Lewis himself dismisses the objection from relevance, writing that «I have often explained what [other cosmoses] have to do with modality, for instance by saying that the modal operators are quantifiers over them». (p. 98) The theory claims that part of what it is for Bush to be a possible loser is for someone resembling Bush to have lost elsewhere. The losses of other people elsewhere are relevant to Bush's possible loss according to the theory because they help constitute it; that's just what the theory says.

This response fails to engage with the constitutional intuition that drives the objection from relevance. I conclude that Lewis thought that constitutional intuitions of this sort should be given little weight in assessing the merits of the theory. The methodological commitment embodied in Lewis's response, then, is that *constitutional intuitions count for little or nothing*.

[20] Philosophers who have offered some version of the objection from relevance include van Inwagen (1985, p. 119), Plantinga (1987, p. 209), and Cameron (2007). It might be what Kripke had in mind by the so-called "Humphrey objection" (Kripke, 1980, p. 45n). It's difficult to tell, since the passage in question is really a joke, and the underlying argument, if there is one, is not explicitly stated.

6. Plenitude and Fit

How are we to assess a reductive hypothesis, if not by appeal to the puzzles it solves or the constitutional intuitions yielded by reflection on its plausibility? One answer offered by Lewis is that we may assess a reductive hypothesis by what we might call *fit*: we ask whether accepting a reductive hypothesis for a certain phenomenon yields a theory which predicts and explains its central features. If it does – if the reductive theory fits the observed features of the phenomenon in question – then that counts in favor of the reductive hypothesis. For instance, by accepting the identification of the temperature of a thermodynamic system with the mean kinetic energy of its molecules, together with some ancillary assumptions, we can derive the ideal gas law from Newtonian mechanics. So, the reductive identification of temperature with mean kinetic energy is supported by its "fit" with observed features of temperature.

Can a similar argument be mounted in favor of Lewis's view? Lewis attempts such an argument in *OPW*. Fit can be achieved by supplementing Lewis's view with claims that guarantee that there is a possible world of the right sort to ground every possibility. To illustrate, suppose that we somehow supplement Lewis's theory so that it yields predictions about what is possible. We then check those predictions against an inventory of the modal facts: that it is possible for there to be purple penguins, but it is not possible for there to be unextended purple penguins, *etc.* Lewis's theory is confirmed to the extent that it predicts a large proportion of the modal facts and contradicts few or none. Lewis in fact holds that there are worlds of the right sort to back every possibility, writing that «absolutely *every* way that a world could possibly be is a way that some world *is*». (p. 2) Thus, Lewis is committed to the truth of every instance of

> (PLENITUDE) If it is possible that φ, then there is a *w* such that *w* is a cosmos and 'φ' is true at *w*

where "true at" for a given sentence φ is analyzed in terms of cosmoses and counterparts. The antecedent of (PLENITUDE) makes a modal claim, and must be analyzed if Lewis is to have a theory containing no modal primitives. Applying a Lewis-style analysis, however, yields a triviality

> If there is a *w* such that *w* is a cosmos and 'φ' is true at *w*, then there is a *w* such that *w* is a cosmos and 'φ' is true at *w*,

which won't help secure a fit between the predictions of Lewis's theory and what's possible. The triviality is satisfied, for instance, even if there is only one cosmos, the one we all inhabit. But such a view would exhibit, given Lewis's analysis of possibility, an extreme lack of fit: it would predict that it is not possible for there to be purple penguins (assuming there aren't any), even though that clearly is possible.[21]

A new idea is needed to secure the right kind of fit between Lewis's theory and the modal facts. Lewis proposes a principle of recombination. The guiding idea is the Humean thought that anything can coexist with anything: possibilities can be combined at will, *modulo* spatial re-arrangements to make sure that everything fits together. For instance, if there could be a nine-foot-tall man and there could be a purple elephant, then a nine-foot-tall man and a purple elephant could coexist, so long perhaps as they occupied different spatiotemporal regions. Lewis employs the notion of a *duplicate* to formulate his principle of recombination. Your duplicate, in this sense, has exactly the same intrinsic properties as you do.[22] Presumably this requires that your duplicate is molecule-for-molecule exactly the same as you, from the skin in. Thus, your duplicate's hair is the same color as yours, her pancreas has the same size, shape, and mass, *etc.* Lewis's principle of recombination is:

> (RECOMBO) If x_1 is an individual in a cosmos w_1, x_2 is an individual in a cosmos w_2, ..., then there is a cosmos w_3 containing any number (including 0) of duplicates of x_1 and any number (including 0) of duplicates of x_2, and ..., size and shape permitting.[23]

(RECOMBO) is not strong enough to achieve the fit Lewis seeks.[24] To the best of my knowledge, no one has ever had skin which had the lime-green color of Oz's Wicked Witch of the West. I might have had skin of that color. But duplicating me and everyone else as many or as few times as you like never yield a person with green skin. If I, everyone else, and our duplicates are the only things that there happen to be that are sufficiently person-like, then

[21] This abbreviated discussion follows and simplifies the discussion at pp. 86–87.
[22] See Lewis, 1983 (pp. 355–361).
[23] See pp. 87–90.
[24] See Divers & Melia, 2002 for an argument that (RECOMBO) requires supplementation so that it says that every spatiotemporal rearrangement of duplicates is realized in some cosmos, size and shape permitting. ((RECOMBO) says nothing about rearrangements.) The present objection applies to such a supplementation of (RECOMBO).

Lewis's recombination principles can be satisfied without supplying a counterpart for me with skin of the right color. (RECOMBO) does not require that there be anything sufficiently person-like other than us and our duplicates. So, (RECOMBO) does not secure fit.

What's more, one might worry that the proviso «size and shape permitting» implicitly smuggles in primitive modality. The most natural way of reading that proviso interprets it as meaning, "so long as it is possible for there to be (in a single cosmos) things of those sizes and shapes, and in that arrangement".[25] If (RECOMBO) is implicitly modal, it inherits the defects of (PLENITUDE) so far as securing the argument from fit is concerned. If it is not implicitly modal, then we need both a non-modal gloss on the proviso, and some reason to be optimistic that this gloss secures the fit we seek.

Lewis's recombination principle does not guarantee that there are cosmoses of the right sort to back every possibility. Lewis himself acknowledges the failure of (RECOMBO) to secure fit: «our principle of recombination falls short of capturing all the plenitude of possibilities» (p. 92). Thus, an attempt to support Lewis's view by appeal to its fit with the observed modal facts does not succeed. Even so, Lewis's embrace of a principle of recombination has been very influential.[26] Also, the methodological doctrine at issue – *reductive hypotheses may be confirmed by fit* – is widely employed.

7. The Significance of *OPW*

OPW's main thesis is audacious, but there are few today who defend it. *OPW*'s lasting significance lies instead in the methods of argument and theory-assessment Lewis developed and deployed on its behalf. I have attempted to illustrate Lewis's methodology by appealing to features of Lewis's exposition and defense of that main thesis. There is much of interest that I have left out, but I hope a clear picture still emerges. According to Lewis, we proceed in theorizing by attempting to pare down the number of primitive notions we need to fully characterize all of reality. We reduce the number of primitive notions by offering analyses, sometimes on a case-by-case basis, of claims

[25] See n. 24 for an explanation of the reference to the arrangement. Thanks to K. Fine and C.S. Jenkins for discussion.

[26] For instance, combinatorialism is taken as axiomatic in Sider, 2007 (p. 52). See Saucedo, forthcoming for helpful discussion and references.

framed using supposedly derivative notions. These analyses will often rely on the existence of unfamiliar entities for their adequacy. We shouldn't worry too much if the requisite existence claims conflict with common sense, even aided and corrected by mature science: the reduction in the number of primitives may compensate for such implausibilities. Reduction of a notion need not be motivated by any special puzzle or problem presented by the facts it may be used to report; having fewer primitives is an independent theoretical virtue in its own right. It is of little or no moment that a reductive hypothesis contradicts constitutional intuitions, so our reductive proposals are not seriously constrained by our intuitions concerning what the reduced phenomenon consists in. We may use considerations of fit to support a reductive hypothesis, or to adjudicate among competing reductive hypotheses. But even a very radical hypothesis may be adopted without the support of considerations of fit, so long as it is sufficiently parsimonious in respect of number of primitives. In general, this methodological orientation strongly favors reduction in the service of securing a smaller primitive ideology. It is a, if not the, dominant methodological orientation of our day.

REFERENCES

Cameron, R. (2007). Lewisian Realism: Epistemology, Methodology, and Circularity. *Synthese, 156*(1),143–159.

Divers, J. (1999). A Genuine Realist Theory of Advanced Modalising. *Mind, 108*(430), 217–239.

Divers, J., & Melia, J. (2002). The Analytic Limit of Genuine Modal Realism. *Mind, 111*(441), 15–36.

Fara, M., & Williamson T. (2005). Counterparts and Actuality. *Mind, 114* (453), 1–30.

Hume, D. (1739/2001). *A Treatise of Human Nature*. Oxford: Oxford University Press.

Kripke, S. (1980). *Naming and Necessity*. Cambridge, MA: Harvard University Press.

Lewis, D. (1968). Counterpart Theory and Quantified Modal Logic. *Journal of Philosophy, 65*, 113–126.

Lewis, D. (1973). *Counterfactuals*. Cambridge, MA: Harvard University Press.

Lewis, D. (1983). New Work for a Theory of Universals. *Australasian Journal of Philosophy, 61*(4), 343–377.

McDaniel, K. (2004). Modal Realism with Overlap. *Australasian Journal of Philosophy, 82*(1), 137–152.

McDaniel, K. (2006). Modal Realisms. In J. Hawthorne (Ed.), *Philosophical Perspectives: Metaphysics, Volume 20*. Oxford: Blackwell, 303–331.

Melia, J. (2008). Ersatz Possible Worlds. In T. Sider, J. Hawthorne, & D. Zimmerman (Eds.), *Contemporary Debates in Metaphysics*, Oxford: Blackwell, 135–151.

Merricks, T. (2003). The End of Counterpart Theory. *Journal of Philosophy, 100*: 521–549.

Nolan, D. (1997). Quantitative Parsimony. *The British Journal for the Philosophy of Science, 48*, 329–343.

Plantinga, A. (1987). Two Concepts of Modality: Modal Realism and Modal Reductionism. In J.-E. Tomberlin (Ed.), *Philosophical Perspectives: Metaphysics, Volume 1*. Atascadero, CA: Ridgeview Publishing Company, 189–231.

Saucedo, R. (forthcoming). Parthood and Location. In D. Zimmerman (Ed.), *Oxford Studies in Metaphysics*. Oxford: Oxford University Press.

Sider, T. (2001). *Four-Dimensionalism: An Ontology of Persistence and Time*. Oxford: Clarendon Press.

Sider, T.. Reductive Theories of Modality. In M. Loux & D. Zimmerman (Eds.), *The Oxford Handbook of Metaphysics*, Oxford: Oxford University Press, 180–208.

Sider, T. (2007). Parthood. *The Philosophical Review, 116*(1), 51–91.

Stalnaker, R. (1976). Possible Worlds. *Noûs, 10*(1), 65–75.

Van Inwagen, P. (1985). Plantinga on Trans-World Identity. In J.E. Tomberlin & P. van Inwagen (Eds.), *Alvin Plantinga*, Dodrecht: Reidel, 101–120.

Commentary

David Lewis
Parts of Classes
Oxford: Blackwell, 1991 *

Einar Duenger Bøhn †
e.d.bohn@ifikk.uio.no

David Lewis's *Parts of Classes* is a great book, in all respects. But one of its most interesting thesis, in my mind, is not its core thesis that standard set theory – ZFC – reduces to classical mereology + plural quantification + a primitive singleton-relation, but rather its sub-thesis of how to understand classical mereology, what Lewis calls the thesis of *Composition as Identity*:

> (CAI): a whole is the same portion of reality as its many parts taken together; it is them collectively, they collectively are it.

CAI is needed as an assumption for the core thesis of *Parts of Classes* – the reduction of ZFC – on pain of it being unmotivated.[1] But CAI is the most interesting in its own right. It is also as such it is presented by Lewis, and received in the literature.[2]

In what follows, I critically assess CAI as Lewis presents it in *Parts of Classes*. I first argue that Lewis's presentation of CAI has been misunderstood in the literature (section 1). I then argue that the best (if not the only) way to understand it entails a slightly revisionary semantics for a certain form of predication (section 2). I finally end by showing that this might create more trouble than it solves for Lewis (section 3).

* Thanks to Giorgio Lando and Roberto Ciuni for comments on an earlier draft.
† IFIKK, University of Oslo.
[1] According to the core thesis of *Parts of Classes*, a class is the fusion of the singletons of its members (and something is a member of a class iff that something's singleton is a part of the class). But why exchange the fusion for the class itself if the fusion is a distinct ontological constituent compared to the singletons of its members? The reduction becomes unmotivated.
[2] The idea behind CAI is not original with Lewis. It is for example proposed by Socrates in Plato's Theateatus (204) and by David Armstrong (1978, 1997). A more radical version of it is defended by Baxter (1988a, 1988b). I defend another version of it in Bohn, ms.

Unfortunately, due to limitations of space, I cannot here go into some recent and most interesting discussions of CAI by especially Ted Sider (2007, ms), Ross Cameron (2007, forthcoming), and Kris McDaniel (2008, 2010).[3] Let it here suffice to critically explicate Lewis's own understanding of the thesis.

1. Composition as Identity Misunderstood

Here is Lewis:

> To be sure, if we accept mereology, we are committed to the existence of all manner of mereological fusions. But given a prior commitment to cats, say, a commitment to cat-fusions is not a *further* commitment. The fusion is nothing over and above the cats that compose it. It just *is* them. They just *are* it. Take them together or take them separately, the cats are the same portion of Reality either way. ... If you draw up an inventory of Reality according to your scheme of things, it would be double counting to list the cats and then also list their fusion. ... I say that composition – the relation of part to whole, or, better, the many-one relation of many parts to their fusion – is like identity. The 'are' of composition is, so to speak, the plural form of the 'is' of identity. Call this the Thesis of *Composition as Identity*. It is in virtue of this thesis that mereology is ontologically innocent: it commits us only to things that are identical, so to speak, to what we were committed to before.[4] (pp. 81–82)

The idea seems clear enough: there is a portion of reality that is ordinarily thought of as some cats, but we can also think of it as one whole thing composed of all and only those cats. That is, in general, the composers collectively and the composed are identical with each other, just thought of under different descriptions.

But Lewis goes on to say that

> mereological relations (however restated) are something special. ... they are strikingly analogous to ordinary identity, the one-one relation that each thing bears to itself and to nothing else. So striking is this analogy that it is

[3] I do so in (Bohn, ms).

[4] Some things xx *compose* something y iff each one of xx is a part of y and each part of y overlaps at least one of xx; x and y *overlap* iff they share a common part; x is a *proper part* of y iff it is a part of y, but is not identical with y; and x is *the fusion of* yy iff yy compose x. *Parthood* is primitive, but reflexive, anti-symmetric, and transitive.

> appropriate to mark it by speaking of mereological relations – the many-one relation of composition, the one-one relations of part to whole and of overlap – as kinds of identity. Ordinary identity is the special, limiting case of identity in the broadened sense. (pp. 84–85)

Lewis then gives five respects in which composition is like ordinary one-one identity, before he concludes that the analogy has its limits:

> In the first place, I know of no way to generalize the definition of ordinary one-one identity in terms of plural quantification. ... And in the second place, even though the many and the one are the same portion of Reality, and the character of that portion is given once and for all whether we take it as many or take it as one, still we do not have a generalized principle of indiscernibility of identicals. It does matter how you slice it – not to the character of what's described, of course, but to the form of the description. What's true of the many is not exactly what's true of the one. After all they are many while it is one. The number of the many is six, as it might be, whereas the number of the fusion is one. And the singletons of the many parts are wholly distinct from the singleton of the one fusion. That is how we can have set theory. (pp. 85–86)

Even though Lewis's initial idea seems clear enough, his subsequent talk of analogy has created some confusion. For example, Peter van Inwagen (1994) takes it to mean that Lewis doesn't really hold that a whole and all its parts are *identical*, but rather that the relation between them is *analogous* to identity. But, as van Inwagen goes on to point out, either a whole and all its parts are identical or they are not identical. If they are merely analogous to being identical, but not really identical, then they are not identical, and the whole is something distinct from its parts, in which case CAI, with its claim that mereology is ontologically innocent, collapses into obscurity.

Beyong-Uk Yi (1999) likewise separates between on the one hand, a *stronger* version of CAI according to which a whole and all its parts are identical literally and strictly speaking, and on the other hand, a *weaker* version of CAI according to which a whole and all its parts are identical only by analogy. Yi, like van Inwagen, interprets Lewis as only accepting the weaker, analogous sense of CAI. (Yi, like van Inwagen, rejects both the stronger and the weaker thesis).[5]

[5] The distinction between a weak and a strong version of CAI is also found in (Sider, 2007), among other places.

But this is a mistake. Lewis proposes what he calls *the thesis of Composition as Identity* according to which a whole (the fusion/the composed) and all its parts (the composers) are one and the same portion of reality thought of under two different descriptions.[6] His talk of analogy is always with respect to the ordinary one-one identity ("x=y"), not with respect to the more general form of many-one identity ("xx=y"). But of course composition is not ordinary one-one identity. After all, composition is a many-one relation, while ordinary one-one identity is not. Recall, according to Lewis, «ordinary identity is the special, limiting case of identity in the broadened sense». So, according to Lewis, there is a more general notion of identity to which composition belongs.[7] So, it is not that composition is analogous to identity as such, but rather that composition is analogous to ordinary one-one identity, which is not saying that composition isn't really identity. Composition is literally and strictly speaking identity in the general sense, but it is not literally and strictly speaking ordinary one-one identity. Lewis's talk of analogy with respect to one-one identity is best thought of as intended to illuminate and motivate a more general notion of identity to which composition belongs.

Much of the reason why van Inwagen and Yi interpret Lewis as only holding the weaker thesis is Lewis's two comments on the limits of the analogy with ordinary one-one identity (quoted above).

The first comment is that there seems to be no way to generalize the definition of ordinary one-one identity in terms of plural quantification:

$x=y =_{df}$ for any zz, x is one of zz iff y is one of zz.

The most natural generalization does not work:

$xx=y =_{df}$ for any zz, xx are among zz iff y is among zz.

Assume my arms, legs, head and torso are identical with my body. Then there is a plurality of things, namely my arms, legs, head and torso, such that my arms, legs, head and torso are among them, but without my body being among them because my body is not one of my arms, legs, head and torso, at least not on any ordinary understanding of 'among', or 'is one of'. So the most natural

[6] Lewis of course allows that the composers are more *objectively natural* – better cut nature at its joints – compared to the composed, or vice versa. See (Lewis 1983, 1986).
[7] See (Lewis, 1993).

generalization does not work. That seems to be Lewis's first point on the limits of his analogy.

But all this shows is that many-one identity cannot be *defined* in terms of plural quantification in the same way one-one identity can. It shows nothing to the effect that composition is not really relating identical things as such. First, plural quantification with ordinary one-one identity doesn't even have the syntactic resources to form well-formed formulas that express many-one identities, so why expect it to be able to define it? Second, identity is a primitive notion if anything is, so the lack of a full definition of it is to be expected. So, Lewis's first point on the limits of his analogy need not and should not be interpreted as a point against composition being identity, but only as a point against composition being ordinary one-one identity.

Lewis's second comment is that there is no generalized principle of indiscernibility of identicals: if xx=y, whatever is true of xx is true of y and vice versa. But if xx is a plurality of more than one thing, then, for example, xx is not one in number, but y is, and xx forms a set that y does not. So there seems to be no generalized principle of indiscernibility of identicals. That is Lewis's second point on the limits of his analogy.

But note, Lewis in *Parts of Classes* also says that

> the many and the one are the same portion of Reality, and the character of that portion is given once and for all whether we take it as many or take it as one ... It does matter how you slice it – not to the character of what's described, of course, but to the form of the description. (p. 87)

So, the point seems to be that there are different ways of describing one and the same thing; one such way is as one whole, another such way is as many parts. According to Lewis, there is no principle of indiscernibility of identicals cutting across, so to speak, all such ways of describing something.[8]

But this does nothing to show that the whole and all its parts are not one and the same portion of reality; it merely shows that some truths are relative to

[8] This same point is perhaps made clearer by Frege (1884, p. 59):

> While looking at one and the same external phenomenon, I can say with equal truth both "It is a copse" and "It is five trees", [...]. Now what changes here from one judgment to the other is neither any individual object, nor the whole, the agglomeration of them, but rather my terminology. But that is itself only a sign that one concept has been substituted for another.

some ways of describing it. So, again, Lewis's second point on the limits of his analogy need not and should not be interpreted as a point against composition being identity, but only as a point against composition being ordinary one-one identity.

2. Composition as Identity Proper

But it is hard – too hard – to understand how xx can be identical with y without xx being indiscernible from y. Contra Lewis, any relation of identity worthy of its name entails indiscernibility. Fortunately, Lewis is unnecessarily pessimistic on this point. Here is a simple suggestion for an appropriately generalized principle of indiscernibility of identicals:[9]

(GPII) $\forall xx \, \forall yy \, (xx = yy \rightarrow (\phi(xx) \leftrightarrow \phi(yy))$,

where xx and yy are plural variables, each taking pluralities of *one or more* things as its value. GPII is just the standard principle of indiscernibility of identicals for ordinary one-one identity, but with plural variables in place for singular variables. To make this an *appropriately* generalized principle, any predication whose truth depends on a form of description of the portion of reality in question must be a *relative* predication, i.e. must be relative to a form of description of the portion of reality it is a predication of.[10] This relative aspect is needed to avoid paradox.

Assume my body is composed of my arms, legs, head and torso, and consider the following three kinds of predication: '...is n in number', '... is one of ...', and '...form set S', where n is a number and S is a set. My body is one in number, but my arms, legs, head and torso are not one in number; my body is one of my body, but my body is not one of my arms, legs, head and torso; and my body forms the set S, but my arms, legs, head and torso do not form the same set S. Now, if, as per CAI, my body is identical with my arms, legs, head and torso, then by GPII we are riddled with paradoxes: one and the same thing both is one in number and is not one in number; it is one of some things and not one of those things; and it forms set S and does not form that set S. But if,

[9] I defend this principle in (Bohn, ms).

[10] The relative aspect of the predication should be placed in the semantics of the predication, not in the syntax, if not only to avoid overcomplicating the notation.

on the other hand, the relevant predications only hold relative to a description of the portion of reality in question, then there are no paradoxes: the portion of reality in question is one in number relative to description D, but not one in number relative to description D*; it is one of some things relative to D, but not one of those things relative to D*; and it forms set S relative to D, but does not form set S relative to D*.

3. Composition as Identity: Trouble for Lewis?

As far as I can see then, given CAI, GPII must be accepted on pain of incoherence, and much predication must be relativized to descriptions for the same reason. But it is not clear that Lewis can accept this, and hence CAI might spell more trouble than it solves for Lewis.

Consider again the above types of predication that need to be relativized in order for GPII to be an appropriately generalized principle of indiscernibility of identicals: '...is n in number', '...is one of...', and '...form set S'. Lewis in *Parts of Classes* takes them all at face value. He treats numerical properties as intrinsic one-place properties of whatever the number holds of; he treats singleton formation as a primitive two-place relation of set theory that holds between a thing and its singleton (see especially ch.1 and section 2.1); and he treats '...is one of...' as a primitive two-place copula of plural logic that relates the singular to the plural (see especially section 3.2). But if there is to be a principle of indiscernibility of identicals along the lines of GPII, as any relation of identity seems to entail, these three types of predication (as well as many others) must be relativized on pain of incoherence. But then, contra Lewis, they are two-place, three-place, and three-place properties, respectively.

REFERENCES

Armstrong, D. (1978). *A Theory of Universals: Universals and Scientific Realism Vol. II.* Cambridge: Cambridge University Press.

Armstrong, D. (1997). *A World of States of Affairs.* Cambridge: Cambridge University Press.

Baxter, D. (1988a). Identity in the Loose and Popular Sense. *Mind, 97*(388), pp. 575–582.

Baxter, D. (1988b). Many-One Identity. *Philosophical Papers, XVII*(3), 193–216.

Bohn, E. D. (ms). On Treating 'Composition' as Identity.

Cameron, R. (2007). The Contingency of Composition. *Philosophical Studies, 136*(1), 99–121.

Cameron, R. (forthcoming). Composition as Identity Doesn't Settle the Special Composition Question. *Philosophy and Phenomenological Research*.

Frege, G. (1884/1980). *The Foundations of Arithmetic*. Oxford: Blackwell.

Lewis, D. (1983). New Work for a Theory of Universals. *Australasian Journal of Philosophy, 61*, 343–377.

Lewis, D. (1986). *On the Plurality of Worlds*. Oxford: Blackwell.

Lewis, D. (1993). Many, but Almost One. In K. Cambell, J. Bacon & L. Reinhardt (eds.), *Ontology, Causality, and Mind: Essays on the Philosophy of D. M. Armstrong*: Cambridge: Cambridge University Press, 23–38.

McDaniel, K. (2008). Against Composition as Identity. *Analysis, 68*(2), 128–133.

McDaniel, K. (2010). Composition as Identity Does Not Entail Universalism. *Erkenntnis, 73*(1), 97–100.

Sider, T. (2007). Parthood. *Philosophical Review , 116*, 51–91.

Sider, T. (ms). Consequences of Collapse.

Van Inwagen, P. (1994). Composition as Identity. *Philosophical Perspectives, 8*, 207–220.

Yi, Byeong-Uk (1999). Is Mereology Ontologically Innocent?. *Philosophical Studies , 93*, 141–160.

Commentary

Andreas Hüttemann
What's Wrong with Microphysicalism?
Routledge, London 2004*

George Darby†
g.a.darby@kent.ac.uk

1. Introduction

In this book Andreas Hüttemann argues against *microphysicalism*, whose «core doctrine is the affirmation of an *ontological priority of the micro-level*» (p. 7). Hüttemann distinguishes three ways of fleshing out the core: *Micro-determination* is the thesis that «the behaviour or the properties of compound systems are determined by the behaviour or the properties of their constituents but not vice versa»; *micro-government* holds that «the laws of the micro-level govern the systems on the macro-levels»; and *micro-causation* «claims that causation takes place in virtue of the causation on the level of the (ultimate) parts» (p. 2). He takes microphysicalism, in these various guises, to be motivated by what he calls *micro-explanation*, «the explanation of the properties on the macro-level on the basis of the properties of the micro-level» (p. 9). His primary concern is then to undermine this move; granting the success of micro-explanation, he argues at length that it nevertheless does not motivate any of those three further theses that might constitute micro-physicalism proper. Along the way various considerations from natural science help to make a compelling case that the doctrine's status as a default assumption in many areas of philosophy ought to be reconsidered.

* I would like to thank Giorgio Lando, Massimiliano Carrara and Roberto Ciuni for the invitation to contribute to this issue, and for valuable comments and discussion, for which I am also grateful to participants at *Hume's Metaphysics and Humean Metaphysics*, Tampere 2011, especially Helen Beebee and Anthony Eagle. The work on this paper was done with the generous support of the Leverhulme Trust.
† Department of Philosophy, University of Kent.

That it *is* a default assumption is reflected in Lewis's famous *Humean supervenience*. At least on the famous statement of the position in the Introduction to the second volume of Lewis's *Philosophical Papers* (1986a), this says that everything is determined by the distribution of intrinsic properties point-by-point. Hence this book's relevance to this special issue, as Humean supervenience appears on Page 1 as a paradigm of micro-determination, Hüttemann citing a portion of that official statement: «all there is to the world is a vast mosaic of local matters of particular fact, just one little thing and then another [...]. And that is all».

To provide further context here, *What's Wrong With Microphysicalism?* touches on a number of themes that converge in recent work. Some of these have to do with arguments for metaphysical holism in quantum mechanics that have been around for a long time. Roughly, the properties of composite quantum systems in entangled states do not supervene on the properties of their component parts, and this has been brought up as a challenge to Humean supervenience by philosophers of science, Maudlin (2007), Ladyman and Ross (2007) being prominent. Similar ideas have recently received attention in mainstream metaphysics – see for example Jonathan Schaffer's recent defence (2010) of monism. Schaffer there also presents essentially philosophical arguments (as opposed to philosophically-interpreted scientific results) for the same conclusion, and this might be seen as exemplifying an increasing anti-reductionism. Since Hüttemann too discusses the use of quantum mechanics in the case against microphysicalism (but with a twist, mentioned below), alongside essentially philosophical arguments, it would not be too much of a stretch to see this book in the same tradition; indeed Schaffer (reviewing this book in Schaffer, 2008) recruits Hüttemann's work in support of his own project.

A further theme, which provides a tempting way of unifying this commentary, concerns the priority (rather than mere relevance) of science in its relation to metaphysics. This is explicitly a theme for those philosophers of science above; less clearly so for Hüttemann. Scientism involves a commitment to engaging seriously with scientific results, which Hüttemann certainly shares, but in the hands of Ladyman and Ross (for example) it goes much further, becoming a polemic against pointless exploration of «entrenched philosophical fetish[es]» (Ladyman & Ross, 2007, p. 21). Some themes in the book chime with the philosophers of science's quarrels with analytic metaphysics, but there is also plenty of engagement with its main concerns, and Hüttemann clearly

doesn't think that the whole enterprise is a waste of time. Nevertheless, I think that there is enough common ground to spin things this way for the purposes of this commentary. Another salient reason will concern causation; mainstream metaphysics in the broadly Lewisian tradition considers homely examples where Billy throws a rock at a window; perhaps Suzy threw one too, which would have broken the window if Billy's hadn't; this kind of example is often used in the literature to draw far-reaching conclusions. Maudlin rejects the Lewisian tradition, preferring what might be called a "laws-first" approach, and Hüttemann does something similar, claiming support from the way in which laws figure in actual science.

I share some of this scientistic conviction too. But, having encountered Lewisian metaphysics at an impressionable age, I would also like to think that it is of more than merely scholastic interest. For this commentary, then, I'd like to think about how Lewisian metaphysicians might reply. Among others, three lines suggest themselves:

1. Deny that the evidence is anything like as conclusive as suggested. This goes especially for quantum mechanics, where the "holistic" reading of the formalism and of Bell's theorem has long been contested.

2. Anti-metaphysical philosophers of science (again acknowledging that Hüttemann might not want to be lumped in with them) tend to engage less with the scholastic details. Well they would, since they think it's a waste of time. But they do offer *rival* accounts of staples of analytic metaphysics; for example (and this certainly does include Hüttemann), alternatives to Lewis's treatment of causation. So we should attack those rival accounts. It would make a nice dialectical point if it could be shown that those scholastic details matter after all, and even nicer if, once the details have been tidied up, the rival accounts tend to converge on something like Lewis's anyway.

3. Dispute more directly the way the dialectical situation has been set up. In this case a strategy would be to wonder in what sense Lewis is really a microphysicalist.

I'll have a go at each in turn in the following three sections.

2. Quantum Mechanics vs Humean Supervenience

Hüttemann presents the standard argument from quantum entanglement, which goes like this (p. 47): If the state of a composite system is a product state, then it can be thought of as being one in which Particle 1 is in state A,

say, and Particle 2 is in state B. But some states, like the singlet state, are not product states. There is no way to think of this as representing Particle 1 in a particular state, without mentioning Particle 2 (and vice versa). Particle 1 and Particle 2 are then said to be *entangled*. Hüttemann puts the lesson like this:

> There are spin states of the compound system such as [the singlet state] that do not allow the attribution of pure states to the parts of the compound. So the fact that the compound is in a determinate state cannot be explained in terms of determinate states the constituents occupy. Here we have an example of the failure of micro-explanation – or at least of one kind of micro-explanation, namely, micro-explanation of the *states* of compound systems.

This marks a departure from the more usual way of talking, which sees the lesson in terms of supervenience or determination: it's not that the compound state is to be *explained* in terms of the states of the parts, so much as that one would hope that it would *supervene* on the states of the parts. Hüttemann does then go on to link this with supervenience, citing Paul Humphreys' discussion (1997). However this is cashed out, though, Hüttemann thinks it somewhat orthogonal to his concerns: entanglement «provides a counter-example to part-whole micro-explanation of states – not, however, a counter-example to part-whole micro-explanation of the dynamics of compound systems» (p. 48). This (the twist mentioned above) leads into a novel argument that, when one pays attention to the *dynamics* of even an entangled quantum system, things appear better for micro-explanation. If I understand right (though Hüttemann doesn't put it this way, hence the qualifier), this would suit Hüttemann's ultimate aim: micro-explanation might succeed even in the quantum case (and see especially the bottom of p. 56), while micro-*determination* (here supervenience) fails. So again, the success of micro-explanation is no evidence for micro-determination, which fits with Hüttemann's overall strategy.

In the present Lewisian context, in any case, what is important is this failure of supervenience. Thus Schaffer, in his review of the book (2008, p. 255) re-emphasises that, whatever the outcome of Hüttemann's dynamic argument, micro-determination is threatened by the loss of «part-whole micro-explanation of states», and Hüttemann too positively endorses this line: «I agree with Humphreys (and others) that quantum entanglement is a case of emergence» (p. 48). So I think Hüttemann can be co-opted into the growing consensus that something like this does pose a very serious challenge to microphysicalism, and Humean supervenience in particular.

So, the key question: How could Lewis respond? One way would be to resist the supposed scientific evidence. Lewis famously hinted at something like this:

> I am not ready to take lessons in ontology from quantum physics as it now is. First I must see how it looks when it is purified of instrumentalist frivolity; [...] doublethinking deviant logic; and [...] supernatural tales about the power of the observant mind to make things jump. If, after all that, it still teaches nonlocality, I shall submit willingly to the best of authority. (Lewis, 1986a, p. xi)

It is often replied that there *are* theories that fit the bill, and they are all nonlocal (they have to be, by Bell's theorem). But it is still open to question what *kind* of nonlocality they involve. A clear refutation of HS requires non-*supervenience*, and the conclusion that the experimental evidence requires *that* may be (and has been, and still is) resisted.

This response, however, leaves hostages to fortune. If the physical evidence, and the consensus (even if not overwhelming) on its interpretation, points to the failure of microphysicalism, then a good metaphysician ought not insist on remaining a microphysicalist. And so Lewis hints at another solution: «[I]f I defend the philosophical tenability of Humean Supervenience, that defence can doubtless be adapted to whatever better supervenience thesis may emerge from better physics» (Lewis, 1994, p. 474). This suggests an alternative line, again putting it in the context both of Hüttemann's book and of the present volume: Lewis thinks that his metaphysics will survive whatever physics throws up. But microphysicalism may well not survive whatever physics throws up; therefore Lewis is not (essentially) a microphysicalist. I'll explore this a bit more in Section 4.

3. FLOTEs and NULA versus Billy and Suzy

Lewis of course is famous for analysing causation in terms of counterfactual dependence, and counterfactuals via comparative similarity of possible worlds. The substantial literature on both of these components is grist to the scientistic mill: Ladyman and Ross (2007, pp. 1, 3, 4, *passim*) see this approach to causation as another folly of analytic metaphysics; McKay Illari, Russo and Williamson's *Causality in the Sciences* (2011) opens with a manifesto along similar lines (at least as far as a more prominent place for science is concerned); and Maudlin (2007) advocates replacing Lewis's analysis of counterfactuals with an approach based on Fundamental Laws of Temporal Evolution

(FLOTEs). Hüttemann too wants to give laws (his account of which gives precedence to their appearance in actual science over the "All Fs are Gs" mould) a more distinctive role, in what he calls the Nomologically Updated Ludovician Account (NULA), his alternative to Lewis's account developed in Chapter 7. The laws-first, science-inspired FLOTE and NULA accounts are ranged against the counterfactuals-first, intuition-guided, tradition of Billy and Suzy.

The kind of reply I have in mind is this: Maudlin's account (2007, Chapter 1) evaluates counterfactuals by fixing things at a Cauchy surface (the relativistic equivalent of a moment in time) so that the antecedent is true, and then letting the laws unfold. This makes it somewhat like the proposal of Jackson's *A Causal Theory of Counterfactuals* (1977); but Jackson's theory faces objections because of this feature (see Bennett, 2003, p. 209). What is interesting in this regard is that Jackson's theory is seen as a step in a series of refinements on a basic idea, which ultimately lead to Lewis's account (Bennett thinks that Lewis's account is not the last word either, but that is a separate issue). Wouldn't it be convenient if the more modern account inherited the same problems, and if the needed fixes were those that motivated the move towards the Lewisian orthodoxy? Well, space sadly precludes exploring this in the case of FLOTEs; but I do think something like this might be pressed against Hüttemann's NULA.

In Lewis's original account (Lewis, 1973a), counterfactual dependence between c and e consists in two counterfactuals ("O" for "Occurs"):

(1) $O(c) \mathrel{\square\!\!\rightarrow} O(e)$
(2) $\neg O(c) \mathrel{\square\!\!\rightarrow} \neg O(e)$

For actual c and e (1) is automatically true and plays little further role in Lewis's account, but an analysis of counterfactual dependence in general might as well include it – the usual possible worlds story (Lewis, 1973b) takes care of their truth conditions. Since Hüttemann wants to do without the possible worlds, and to avoid the mess of what to put in their place, he simply drops (1), swapping it (pp. 110, 112) for

(1*) $O(c) \wedge O(e)$

The major departure from Lewis's system appears to be with (2), whose truth-value is to be «entirely a matter of laws of nature». Of course laws *figure* in Lewis's analysis: $p \mathrel{\square\!\!\rightarrow} q$ is (actually) true iff (more-or-less) the closest p-world

to actuality is a q-world, and comparative closeness to the actual world depends on match of particular fact and the extent of violations of the actual laws (Lewis, 1979). But the laws themselves emerge from the Humean mosaic via the Best System analysis.

Hüttemann's alternative approach is illustrated by two colliding billiard balls, A and B. A causes B to change direction; for concreteness let's say that A caused B to miss the pocket; the specific counterfactual he considers is *if A had not collided with B, B would have taken path b* rather than path b* (so b* in our case leads to the pocket, b does not).

The Lewisian analysis would go something like this: the minimal change (on the Lewis, 1979 criteria) has perfect match until some small violation of the actual laws to prevent the collision, thereafter everything unfolds as the laws require and B goes in the pocket. That gives the counterfactual dependence *if A hadn't collided with B then B would not have missed the pocket*, and the further analysis of causation in terms of that dependence delivers the required result – that the collision caused the miss. At this point the Lewisian analysis famously has a problem with pre-emption: take a situation where a bee was flying across the table and would have knocked B away from the pocket if A hadn't got there first. Now *if A hadn't collided with B then B would not have missed* is false, even though we still want to say that the collision with A caused the miss. This appears to be an ongoing glitch in the program, so it would be handy if Hüttemann's analysis avoided it.

So, the crucial passage is:

> (2) turns out to be true because there is a law that tells us that B would have continued along path b* if it had continued to be isolated. If B were isolated it would behave according to the Hamilton equations with the Hamilton function $H = p^2/2m$. Less pretentiously, it is Newton's first law that tells us how B will continue in the absence of a collision. The counterfactual (2) is true because there are laws about what would happen in the absence of the cause-event.
> This analysis is adequate in general. We do not need possible world semantics for (1*) and (2). (1*) simply registers that the cause-event and the effect-event have occurred. (2) cites a law that states how a system that goes into the effect-event would have behaved if it had remained isolated. (p. 113)

The idea, then, is that the required counterfactual is made true by B's velocity shortly *before* the collision, plus Newton I, which has the isolated B travel straight to the pocket.

Hüttemann sees this as enjoying advantages over Lewis's account, importantly that it avoids the problem of pre-emption. Since he stipulates that the relevant counterfactual consider B in *isolation*, not just from A but from all outside interference, including the bee, on his understanding *if A hadn't collided with B then B would not have missed the pocket* will come out *true*, as required. So too, then, the NULA account of causation has it come out true that the collision with A caused B to miss the pocket.

So far so good. However:

(a) In Chapter 2, and starting with Gallileo on free-fall, Hüttemann develops an unHumean account of laws, drawing on the idea of *continuously manifestable dispositions* that may be partially manifested in real situations, to be contrasted with dispositions whose manifestation is an «all or nothing affair» (p. 19). This plays a crucial role in Chapter 7, with the thesis that

> laws of nature describe how systems would behave if they were isolated. [...] According to NULA it is exactly these kinds of counterfactuals that are appealed to in condition (2). [...] Laws of nature tell us how a system would behave if its behaviour were not caused to change by some external factor. (p.113)

But B in *complete* isolation, though it might indeed continue in a state of uniform motion (assuming there is then enough spacetime structure for this), won't reach the pocket (since there is no pocket to reach). The system needs to include at least the table, but not A, or the bee. «The absence of the cause-event» and «in the absence of a collision» suggest something more like this, but «if it had *remained* isolated» suggests something else again. So I suspect that there is value in objecting that more information is needed on what is meant by «the system», and «isolated». (On a similar theme, contrary to the quote above, neither (2) nor its particular instance in the case at hand appears to *cite* anything so general as a law at all.)

(b) Next, given a complete account of the relevant understanding of the *system in isolation*, we can look for counterexamples. Suppose I twirl a conker on string. The string's breaking causes the conker to fly off towards your window. Here Hüttemann needs a counterfactual something like

> If the string hadn't broken, the conker would have continued to travel (approximately) in a circle).

But there doesn't seem to be any way of achieving an unbroken string simply by stipulating that the system remain isolated – in this case it's not, as suggested

in the quote, a matter of an *external* factor that needs to be excluded (that's what isolation achieves), but an *internal* factor that needs to be altered. (How does it need to be altered? Well, the state of the world at the time of the antecedent could simply be changed to make the antecedent true – the string suddenly breaks – as in Jackson's account. But the problems with Jackson's account (to do with sudden changes when large cause-events are considered) motivate a revision to having a small miracle some time before neatly bring about the antecedent, ... and we are back to Lewis.)

(c) Pre-emption problems are avoided because so much is excluded from the relevant law. But that also introduces spurious causes: Hüttemann is explicit that two events *c* and *e* causally depend on one another if (1*) and (2) hold (p. 114), so it appears that this is a sufficient condition for causation. That means that any example of actual *c* and *e* that fits Hüttemann's way of cashing out (2) will give an example of causation. On that account, what ensures that the collision with A gets correctly attributed as a cause of B's missing the pocket is the fact that the law applied to just the system composing B (and the table), *in the absence of the collision with A*, results in B not missing the pocket. But as far as absence from the relevant system including B *in isolation* is concerned, there was nothing special about A. The same law, applied to the system composing B (and the table) in isolation, *in the absence of more-or-less any event you choose*, results in B taking path b* to the pocket. That seems to ensure that more-or-less any event gets attributed as a cause too. Assuming that this result is undesirable, that would show that NULA is too liberal.

Perhaps what is needed is an additional clause concerning what happens when A is considered after all – but that would be something like Lewis's original clause (1) (but with Hüttemann's laws-based approach to its evaluation). Alternatively, the specification of what it is to consider B *in isolation from A* could be tweaked, perhaps to reflect the *minimal* change needed to bring about ¬A. But again, that is what Lewis's original analysis does.

(d) Lewis's analysis proceeds by focussing on one particular counterfactual:

> If c hadn't happened, e wouldn't have happened.

That counterfactual, like any counterfactual, has its truth-value already, perhaps, as for Lewis, fixed by the general theory of counterfactuals (and note that Lewis's theory is built to deliver the truth-values that we all agree are correct – it is not prescriptive). This particular counterfactual may turn out to

be unsuitable for the analysis of causation: it clearly has one truth-value, Lewis needs it to have the other (as in the pre-emption problem). One strategy would be to find a different counterfactual, or perhaps to introduce more machinery such as *quasi-dependence* (as in the postscripts to *Causation* added in Lewis, 1986b, pp. 205ff). Whatever the details, the change is being made to the theory of causation, not the theory of counterfactuals. But Hüttemann's response *does* change the truth-value of the relevant counterfactual, to suit his specific purpose. This seems to change the wrong component. In the pre-emption case, if A and B hadn't collided, the miss still *would* have happened. That is *why* the counterfactual

If A hadn't collided with B then B would not have missed the pocket

comes out false, contrary to Lewis's requirement that it be true. The theory of causation should be revised to cope with the truth-values that counterfactuals are agreed to have, not the other way round.

I think there is mileage here in defence of the Lewisian approach by pressing these kinds of objections. Moreover, I would hope that tweaking Hüttemann's account to avoid the objections may take it back towards Lewis's. This reflects a general strategy in defence of analytic metaphysics: show that its products, while perhaps not drawn from the latest science, are of use even to its critics. (A similar strategy is suggested by Katherine Hawley, in her review (2010) of Ladyman & Ross, 2007, where she notes that Ladyman and Ross could use some of those products to overcome problems in their own brand of metaphysics.)

4. In What Sense is Lewis a Microphysicalist?

Microphysicalism is certainly questionable from a variety of directions, as Hüttemann shows. I've suggested above some ways in which one might reply to some specific details in respect of Lewisian metaphysics, but (except insofar as the details are components in Hüttemann's overall plan) nothing to defend microphysicalism itself. Perhaps it can be defended (undermining the argument from entanglement would be a part of this), but suppose not. Then Lewis will have to make some kind of concession on the details. But perhaps those details (in particular, as suggested at the end of Section 2, the microphysicalism) are disposable.

For example, in the context of the conflict with quantum mechanics, I think that once one pays attention to the distinction between the *letter* (the official statement from Lewis, 1986a cited by Hüttemann as a paradigm of microphysicalism) and the *spirit* (the denial of necessary connections between distinct existences) of Humean supervenience, there is no real problem with entanglement – whether reality consists of many particulars instantiating spatiotemporal relations, or whether it consists of many particulars instantiating some other kind of relations, doesn't really matter to the core project. Whereas Hüttemann, then, sees microphysicalism as being driven by an argument from the success of micro-explanation, and no doubt that is the case for some of its defenders, for Lewis the microphysicalist component of Humean supervenience is merely a working assumption, and requires no particular motivation beyond a received view of classical physics.

I am fairly sure that this is the natural line of defence against Ladyman and Ross *et al.*, (and have argued for it in Darby, 2009). As further evidence for the spirit/letter, or perhaps hard core/protective belt, distinction, note that Borghini and Lando in this volume (their Section 4) find it natural to identify a Weak and a Strong version of Humean supervenience along similar lines. Like Borghini and Lando, I think that the thesis, among others, that the fundamental relations are the spatiotemporal ones is «*a posteriori* and concerns only our world and other worlds sufficiently similar to ours» (p. x, this volume). In the remainder of this essay I would like to acknowledge some doubts about this.

Famously Lewis says that HS is *contingent* in the passage immediately following the official statement of the doctrine in the introduction to the second volume of his *Philosophical Papers* (1986a, p. x):

> Two worlds might indeed differ only in unHumean ways, if one or both of them is a world where Humean supervenience fails. Perhaps there might be extra, irreducible external relations, besides the spatiotemporal ones; there might be emergent natural properties of more-than-point-sized things; there might be things that endues identically through time or space, and trace out loci that cut across all lines of qualitative continuity. It is not, alas, unintelligible that there might be suchlike rubbish. Some worlds have it. And when they do, it can make differences between worlds even if they match perfectly in their arrangements of qualities.

Here Lewis appears to endorse an inference from conceivability to possibility («It is not, alas, unintelligible [...] [Therefore?] Some worlds have it.»). But elsewhere, for example in Lewis, 1986b (p. 90), he explicitly rejects

this. Again, in Lewis, 1983 (p. 362), he feels the need to allow for the possibility of epiphenomenal spirits, and the consequent need to formulate a contingent supervenience thesis is one of the motivations for natural properties in his metaphysics; but in Lewis (1986b, p. 73), wondering how spirits might be part of a world, given that worlds are unified by spatiotemporal relations, and having surveyed some ways round, such as allowing them to be related in time, he concludes: «I am not sure why I need to defend the possibility of spirit tales – after all, people have been known to accept impossible theories, as witness naive set theory – but in fact I think I give them at least as much room in logical space as they deserve». In the standard presentation, at least, the unification of worlds by spatiotemporal relations, and so the absence of disembodied spirits, are not contingent, and Lewis is content with this. Nor, partly because of the way modal realism is set up, is the denial of necessary connections. In a similar vein, the denial of dualism, which is surely no mere working assumption, might be guaranteed by locality: if there were irreducible mental properties then these would require more than a point at which to be instantiated (thanks to Helen Beebee for suggesting this).

The conflict with quantum mechanics, and the importance of microphysicalism in general, hinges partly on the sense in which Humean supervenience is contingent, and especially the status of spatiotemporal relations in this regard. There appears to be some flexibility in interpretation about why exactly Lewis's supervenience thesis *has* to be contingent, and how strong that commitment is. Clarifying this will give a better handle on the consequences of arguments such as Hüttemann's that appear to show that the doctrine is, as standardly presented, false.

5. Conclusion

Hüttemann's book contains compelling reasons to re-think, if not finally reject, the microphysicalist assumptions common to much of contemporary philosophy. This situates it in a growing and welcome movement towards more scientifically informed metaphysics. But the details matter both ways, and those details offer plenty of scope for reply by Lewisian metaphysicians. I look forward to seeing how this pans out in the near future.

REFERENCES

Bennett, J. (2003). *A Philosophical Guide to Conditionals.* Oxford: Oxford University Press.

Darby, G. (2009). Lewis's Worldmate Relation and the Apparent Failure of Humean Supervenience. *Dialectica , 63*(2), 195–204.

Hawley, K. (2010). Critical Notice of *Every Thing Must Go. Metascience, 19*(2), 174–179.

Humphrey, P. (1997). How Properties Emerge. *Philosophy of Science, 64*(1), 1–17.

Jackson, F. (1977). A Causal Theory of Counterfactuals. *Australasian Journal of Philosophy, 55*(1), 3–21.

Ladyman, J., & Ross, D. (2007). *Every Thing Must Go.* Oxford: Oxford University Press.

Lewis, D. (1973a). Causation. *Journal of Philosophy, 70*(17), 156–163.

Lewis, D. (1973b). *Counterfactuals.* Oxford, Blackwell.

Lewis, D. (1979). Counterfactual Dependence and Time's Arrow. *Noûs, 13*(4), 455–476.

Lewis, D. (1983). New Work for a Theory of Universals. *Australasian Journal of Philosophy, 61*, 343–377.

Lewis, D. (1986a). *Philosophical Papers, Vol. II.* Oxford: Blackwell.

Lewis, D. (1986b). *On the Plurality of Worlds.* Oxford: Blackwell.

Lewis, D. (1994). Humean Supervenience Debugged. *Mind, 103*(412), 473–490.

Maudlin, T. (2007). *The Metaphysics within Physics.* Oxford: Oxford University Press.

McKay Illari, P., Russo, F., & Williamson, J. (2011), *Causality in the Sciences.* Oxford: Oxford University Press.

Schaffer, J. (2008). Review of *What's Wrong with Microphysicalism? British Journal for the Philosophy of Science, 59*(2), 253–257.

Schaffer, J. (2010). Monism: The priority of the Whole. *Philosophical Review, 119*(1), 31–76.

Commentary

The Universe as an Ongoing Enterprise. Commentary on *The Metaphysics Within Physics* by Tim Maudlin

Emiliano Boccardi[†]
emiliano.boccardi@gmail.com

Tim Maudlin's book *The Metaphysics within Physics* (Oxford University Press, 2007, since now on 'MWP') is a collection of six essays written over a span of over a decade. Although thought of as independent and self-contained, they clearly constitute a consistent, unified proposal for a physics-based ontology. Many before Maudlin thought that ontology should be informed by, if not derived from, physics. What is peculiar with his proposal is not so much that physics inspires his ontology, but that in deriving metaphysical lessons from physics, he is not driven by empiricist scruples. Indeed, the overarching philosophical setting of the proposal, as well as the dialectical structure of its narrative, are driven by anti-empiricist sentiments. We live in a time when empiricist theories of meaning are a relic of the past, thinks Maudlin. Verificationist intuitions have long been gone. We are now allowed again to refer to the non-linguistic items that we think populate our world, just like logicians refer to the models of their formal systems. At any rate physicists, according to Maudlin, do. And he thinks ontologists should do the same.

On the notes of this *de profundis* of empiricist theories of meaning, Maudlin proceeds to attack the philosophical heuristics that has always been associated with them. Notably, he questions the epistemic value of Okkam's razor. *Entia non sunt multiplicanda praeter necessitatem*, we have been told many times. But what should we take "necessity" to mean in these cases? Certainly not *logical* necessity for, argues the author, in that case the razor would "lend us in solipsism". On the other hand, if we understand the dictum as suggesting that one should not inflate one's ontology *without good reason*, then «the principle becomes a harmless bromide: nor should one's ontology be

[†] University of Geneva and University of Fribourg, Switzerland.

reduced without good reason» (MWP, p. 3). The reader is now ready for some rationalist, physics-based proliferation of entities. Before introducing and discussing Maudlin's world, let me introduce the (familiar) desert landscape where it is built. Most of the arguments presented in the book are aimed at debunking Humean Supervenience: «the doctrine that all there is to the world is a vast mosaic of local matters of fact, just a little thing and then another. [...] For short: we have an arrangement of qualities. [...] All else supervenes on that» (Lewis 1986, p. x). Such arrangement of properties of spacetime points and elementary bits of matter (together with the external relations that they instantiate) constitute our deserted landscape: the "Humean mosaic". Lewis, and so many others, believe that this Mosaic is all there is to the world.

The strategy of Maudlin's attack on Humean Supervenience is based on an analysis of the thesis into two sub-theses: Separability and Physical Statism. Put roughly, Separability is the thesis that all fundamental properties are local and that the only fundamental external relations are spatiotemporal relations. Physical Statism is the thesis that «all facts about a world, including modal and nomological facts, are determined by its total physical state» (MWP, p. 51). The conjunction of these theses, according to Maudlin, is the doctrine of Humean Supervenicence. The practice and the believes of contemporary physicists – this is the main negative thesis argued for in the book – is incompatible with both Separability and Physical Statism. Separability is argued to be incompatible with Quantum Mechanics (QM). Physical Statism, instead, with common epistemic and explanatory scientific practices. The first lesson that ontologists should learn from physics is negative: Humean Supervenience is false.

We shall take up Maudlin's negative arguments in turn. Before doing so, however, let me briefly sketch the positive theses. Laws of nature, according to Humean Supervenience, are nothing but patterns of the Mosaic. The Mosaic just happens to instantiate certain patterns: we don't need to add these patterns to our ontology, for the Humean Mosaic is more than enough to instantiate them. Most analyses of the concept of law assume that their logical form is that of a universal generalization: $(x)(Fx \rightarrow Gx)$.This formulation had a lot of appeal to the logical empiricists, as it allowed them to specify the syntactic properties of laws. But, once lifted from the burden of having to capture the syntactic structure of scientific reasoning, one is no longer sure that laws are (essentially) universal generalizations. Maudlin's anti-empiricist package moves from the observation that laws as they are sought (and sometimes found)

by physicists don't appear to have (essentially) that form. Think of the fundamental law of Newtonian Mechanics:

$$F = \frac{dp}{dt}$$

Or of the fundamental law of (non relativistic) Mechanics:

$$ih\frac{\partial}{\partial t}|\psi >= H|\psi >$$

These mathematical formulations, observes the author, are aimed at describing how the state of a system evolves through time. To be sure (MWP, p. 11), they can be stretched into the form of universal generalizations, but this is neither necessary nor obviously perspicuous. I think Maudlin is right in claiming that universality is not (nor should be) explicitly inbuilt into laws. It makes perfect sense, for example, to suppose that a law governs the behavior of matter in a region of the universe, but not in others, or that it governs it at certain times but not others. This observation, by itself, doesn't constitute an objection to the Humean account of laws. Laws that only apply to restricted portions of spacetime may nonetheless be viewed as restricted, parochial generalizations. But virtually all truths (even singular ones) can be tortured into the form of more or less parochial generalizations. What is at issue here, however, is whether the mere fact that a truth can be expressed by a generalization suffices to characterize it as a law of nature. Maudlin thinks not. What characterizes laws, he thinks, is their role in explaining how certain states are *produced* from previous ones:

> [when] providing explanations and accounts of things, what we actually do is specify the state of the initial (i.e. earliest) boundary, and regard the state interior and on the final boundary to be explained or produced from the initial conditions and the operation of the laws through time. (MWP, p. 130).

Ned Hall (2004) distinguishes two concepts of causation: dependence (mere counterfactual dependence) and production (a stronger notion of cause that accounts for the production of later states from previous ones). Drawing heavily from this distinction (MWP, p. 174), Maudlin claims that laws should explain how subsequent states are *produced* from earlier ones, and not merely how they counterfactually *depend* on one another. As we shall see, according to Maudlin laws can fulfill this conceptual role only if (1) they are taken to be irreducible, and (2) not supervenient on the Humean Mosaic; and, finally, (3) if

the passage of time is taken to be an irreducible fact about the spacetime manifold.

Summing up, once one has (1) the laws, (2) the initial state of the world and (3) once the whole thing is "set in motion" by introducing the passage of time, the rest of the Mosaic gets produced as time unfolds. All other facts about it are thereby determined and productively explained.

> The universe started out in some particular initial state. The laws of temporal evolution operate, whether deterministically or stochastically, from that initial state to generate or produce later states. And the sum total of all the states so produced is the Humean Mosaic. This counts as an explanation exactly because the *explanans* (namely the initial state, or the state up to some time, and the laws) are ontologically distinct from the *explanandum* (namely the rest of the mosaic). (MWP, p. 174–175)

Let us now go back to the negative arguments against Humean Supervenience. As we said, according to Maudlin, Separability, the doctrine that «the complete physical state of the world [...] supervenes on the intrinsic physical state of each spacetime point» (MWP, p. 51), is incompatible with Quantum Mechanics. The arguments to this effect are proposed at the beginning of the second essay. The phenomenon that causes troubles for Separability is the so called phenomenon of quantum entanglement. Consider a system comprising two electrons. For any state in which the first electron can be and any state in which the second one can be, there is a state of the whole system that assigns the first state to the first particle, and the second state to the second one. These are called *product states*. For example, the first electron may be in z-spin state up ($|z\uparrow\rangle_1$), and the second in z-spin state down ($|z\downarrow\rangle_2$). The joint system, in this case, will be found in the product state $|z\uparrow\rangle_1 \, |z\downarrow\rangle_2$. Product states are separable, in that they assign a determinate state to each component particle: they consist of the logical sum of these individual states. The trouble for Separability is caused by the so called *principle of superposition*. If A represents a quantum state of a system, and B another, the principle affirms that a third possible state of the system is represented by $\alpha A + \beta B$, where α and β are any two complex numbers such that $|\alpha|^2 + |\beta|^2 = 1$ (this complex state is called a *superposition* of A and B). A consequence of this principle is that any possible pair of states for the whole system can be superimposed to yield a new possible state.

Consider, for example, the two product states: $|z\uparrow\rangle_1 |z\downarrow\rangle_2$ and $|z\downarrow\rangle_1 |z\uparrow\rangle_2$. The principle of superposition affirms that the following are also possible states for the system:

Singlet State: $\frac{1}{\sqrt{2}}|z\uparrow\rangle_1 |z\downarrow\rangle_2 - \frac{1}{\sqrt{2}}|z\downarrow\rangle_1 |z\uparrow\rangle_2$

Triplet State (m=0): $\frac{1}{\sqrt{2}}|z\uparrow\rangle_1 |z\downarrow\rangle_2 + \frac{1}{\sqrt{2}}|z\downarrow\rangle_1 |z\uparrow\rangle_2$

Notice that these states cannot be expressed as products of pure states of each particle: no *pure* quantum state for each single particle yields the same statistics as those of a system in the singlet or in the triplet state. The state of the whole system, in these cases, does not appear to supervene on the individual states of its components. This is why, *prima facie*, these states (also known as *entangled* states) constitute a threat to Separability. One may be tempted to salvage Separability by interpreting entangled states as statistical *mixtures* of pure states. According to this interpretation, each particle is really in a determined pure state, but there is epistemic uncertainty as to which one it is. This fits well with the statistics that one observes when performing measurements on each individual particle. Performing a z-spin measurement on the first electron in a singlet or in a triplet state, for example, one has 50% chances of finding it in z-spin state up ($|z\uparrow\rangle_1$) and 50% chances of finding it in z-spin state down ($|z\downarrow\rangle_1$). This is exactly what one would observe if the particle were in an (impure) statistical mixture of $|z\uparrow\rangle_1$ and $|z\downarrow\rangle_1$. In other words, as far as single particle measurements are concerned, entangled states are not incompatible with the single particles being in a determined pure state: one needs only assume that there is (epistemic) uncertainty as to which state each particle is in.

This maneuver, however, does not allow us to salvage Separability. As noted by Maudlin (p. 59), in fact, spin measurements of a single particle will not allow us to detect any difference between a system in the singlet and one in the triplet state: in both cases, each particle has a 50% chance of being found in z-spin state up and a 50% chance of being found in z-spin state down. So, if Separability held, we should conclude that pairs of particles in the singlet state are in the same state as pairs in the triplet state, contrary to the different statistics that we observe when we perform global, rather than local, measurements. So, argues the author, if QM is true, Separability fails.

This argument strikes me as problematic, for at least two reasons. For one, Maudlin thinks that the argument is sound under any (sensible) interpretation

of QM: «the upshot is that no physical theory that takes the wave function seriously can be a separable theory» (MWP, p. 61). If QM is (part of) a true description of the world, then the world is not separable. The validity of this argument, it seems to me, trades on a confusion between the notions of Separability and Locality (the thesis that the state of a system can only be directly influenced by its immediate surroundings). There exist models other than QM that are Separable (albeit non local) and that replicate the predictions of QM (models that include backward in time connections or non local correlations outside the light cone are some examples). So the argument should be made conditional upon rejecting these (admittedly exotic) models. This, however, would substantially reduce the strength of the argument. Wasn't it precisely because QM appears to violate Separability that so many physicists (notably Einstein) strived to "complete" the theory? Shouldn't we take seriously the intuitions and epistemic practices of so many physicists?

Sure these exotic models are not main stream, but they certainly are part of physics just as QM is. Moreover, very few believe that QM is the last word on microphysics. As Lewis put it: «I'm not ready to take lessons from Quantum Physics as it now is. First I must see how it looks when purified of instrumentalist frivolity, and dares to say something about [...] the constitution of the world» (Lewis 1986, p. xi). The author takes notice of these models (MWP, p. 62, note 5) but, I think, fails to tell us how we are to resolve this underdetermination. Responding to the quoted skeptical remark by Lewis, Maudlin says that «[o]ne *can* see how quantum mechanics looks like when purified of instrumentalism, and quantum logic, and consciousness-induced wave collapse» (MWP, p. 63). He has in mind theories such as David Bohm's so called ontological interpretation, or the Ghirardi-Rimini spontaneous collapse model. Thus "purified", thinks the author, these theories «all agree that the physical state of the world is non separable» (MWP, p. 63).

Maudlin, however, fails to notice that in Bohmian mechanics the wave function "lives" in configuration space, not in 3-D space. According to this interpretation, any particle is always found in an *actual* configuration, representing its (complete) state in a determinate, objective way (even in circumstances when nobody observes it). It is thus clearly separable (albeit non local). The wave function, and the configuration space where it evolves according to Schrödinger's equation, by contrast, cannot be interpreted as describing the objective state of the particle, or even the possible objective states in which the particle would be found if a measurement were to be

performed. What Separability amounts to in configuration space, if one adopts a Bohmian interpretation, is still a highly controversial issue. It is controversial precisely because it certainly isn't the same as in the case of the space of actual configurations (see Suárez 2009 for a discussion of this point). Although this may appear to be a minor technical quibble, it serves the purpose of showing how the issue of Separability cannot be unambiguously settled by considerations that pertain to the practice of physics alone.

Let us now turn to the arguments against Physical Statism. As we said, these arguments aim at showing that the explanatory practices of physicists are not compatible with Physical Statism. The validity of these arguments, in this case, is not conditional on the acceptance of any particular theory, hence their scope and generality is much greater. If laws of nature are to fulfill their explanatory role, it is argued, it must be possible for different laws to generate the same physical state. Different laws, in other words, may share some or all of their models. The paradigmatic example is the theory of general relativity (GTR). According to Maudlin two otherwise identical worlds may differ solely in that one is governed by GTR and the other by the special theory of relativity (STR). GTR and STR share at least a model: empty space. If Physical Statism were true, reasons the author, then any model (the Mosaic) should suffice to determine what laws govern it. But this, we have seen, is not the case, at least in our example. So Physical Statism must be false.

In a recent review, Mauricio Suárez (2009) claims that this argument suffers from a fatal circularity. He argues that one has to already have a *robust* conception of laws (i.e. one according to which laws are ontologically independent from the Mosaic that they produce) to find the argument compelling. The Lewisian, he claims, would simply not concede that both STR and GTR could feature in the best system laws, hence she would not concede that they provide us with an example of different laws sharing the same model. The Lewisian would claim, for example, that the laws of STR constitute at best approximately true regularities, so they do not count as genuine laws to begin with. I think this criticism is unfair. Maudlin's contention that different systems of laws may produce the same Mosaic (i.e. have the same model) is not based on having already endorsed a *robust* account of laws: it derives from the desideratum that laws be compatible with standard explanatory practices and intuitions. We think that what explains the motion of the moon over a period of a year is the initial disposition of matter plus the laws of GTR. Of course, if instead of being governed by the laws of GTR, this portion of the world were

governed by some stochastic laws according to which any position of the moon at each instant during this year were equally likely (given the previous ones), then the explanatory force of the initial position of the moon would be bankrupt. Yet the resulting Mosaic would be just the same. So, if the initial conditions are to have any explanatory role in case the world is governed by the laws of GTR, then the laws themselves must be ontologically independent from the Mosaic, for the *explanandum*, qua *explanandum*, must be ontologically independent from the *explanans* (in this case the Mosaic itself).

The metaphor of the mosaic is quite apt to expose this reasoning. Suppose you enter a temple and find yourself in front of a large mosaic, apparently representing Plato surrounded by his disciples. As you look at it, you ask yourself why the mosaic represents Plato as wearing a baseball cap. Someone tells you that the mosaic was produced by a random procedure whereby each piece was positioned according to a randomizing algorithm. This information allows you to explain what you are seeing. What the mosaic represents doesn't need an explanation. The patterns you are watching do not even contain a representation of Plato: the mosaic arguably doesn't represent anything. Suppose further that later you discover that the information was false: really, the Mosaic was the winner of a competition for young artists. You are told what the intentions of the artist were. Again, this provides you with a good (albeit different) explanation. Yet you are looking at the same mosaic! The features of the same mosaic may have different explanations, or not have any, depending on what rules, if any, governed its production. Clearly the rules, if there are any, must not supervene on the mosaic itself, if they are to have any explanatory teeth. Now, Maudlin's metaphysics is not meant to account for human intentions, so his arguments do not literally apply to our example. But I don't see why the reasoning should lose any of its force if, instead of the human intentions, what explained the mosaic was some other, non random algorithm, together with the details if its implementation.

The case is even more striking if we consider different statistical laws that produce the same models. Consider a coin that is tossed ten times to yield ten heads. Is it a fair coin, or is it biased? Sure it's more likely to be biased, but the mere possibility of it being fair opens the way to different sets of laws that produce the same mosaic. In one scenario, for example, it is a law that the chances of a coin like the one that was tossed landing head is 50% (notice that the probability is built into the law). In a different scenario, it is a (statistical) law that the chances of landing head are only 10%. Any sequence of heads,

including our sequence of ten heads, is a model of both laws, so, it is argued, such laws cannot supervene on any particular sequence that the Mosaic happens to instantiate. Also this line of reasoning has been argued to be circular. It assumes, rather than showing, that chances consist of chance ascriptions based on stochastic laws. But what if chances inhere objectively in the events themselves (for example by being instantiated by some intrinsic dispositional properties of the events)? In this case chances would be primitive (relative to the laws): they would belong to the mosaic, rather than being imposed on it by the operation of laws. The laws would just serve the purpose of describing the evolution of such chances through time.

Again, I think this criticism is unfair. True, the arguments that we have discussed do not force us, by themselves, to buy Maudlin's package. If they are sound, these arguments only prove that, if we take the explanatory practices of physics at face value, laws must be thought of as ontologically independent on the mosaic. However, they do not suffice to prove that we should take them as primitives of our ontology. But Maudlin's claim is not to have provided compelling reasons in favor of his package. More modestly, he aimed at providing a metaphysics that is as faithful as possible to the unreflective believes and practices of physicists. Physicists don't analyze laws of nature but content themselves with seeking them and using them in their productive explanations. Well, then unless we have good reasons to think otherwise, let's stick to the idea that they are primitive. Physicists typically try to explain the observed patterns of the Mosaic. Then, unless we have reasons to think they are mistaken, let's take these patterns as legitimate *explananda*.

This is no knock down argument against the Humean picture. You can't ask the Lewisian to explain a regularity in the pattern, or accuse him of not being able to explain it. Lewis is not claiming that the patterns of the Mosaic are produced by random processes. No, to the mind of the Lewisian the patterns are just not produced at all: they are what they are and the need to explain them is but a human illusion. To this, Maudlin can only retort that these are not the intuitions of practicing physicists. That's all.

> The first order of business has been to show that my non-Humean package really is an alternative account that runs into no obvious logical, methodological, or scientific objections. [...] The metaphysics within physics is, as of now, non-Humean, and we can do no better as philosophers than embrace it. (MWP, p. 182)

Let me now complete the presentation of the non-Humean package. As we said, the primitives of Maudlin's ontology are the laws of nature and the passage of time. We have discussed the reasons for thinking that laws are primitive. It is now time to discuss the role of the passage of time in Maudlin's package. The passage of time serves primarily the purpose of doing justice to the intuition that the world is "made" as time goes by, that it is an "ongoing enterprise". Remember that, according to Maudlin, physicists seek *productive* explanations: «the universe, as well as the smaller parts of it, [...] is *generated* from a beginning and guided towards its future by physical law» (ibidem). The fourth essay of Maudlin's book is entirely devoted to defending the thesis that the passage of time is real and irreducible. What is peculiar with Maudlin's understanding of this thesis is that he is arguably a static eternalist: he believes that all events, past, present and future, are all equally real (the ontological status of an event in no ways depends on its temporal location) and that presentness, pastness and futurity are not absolute, monadic properties of times (or events). It is worth quoting at length how Maudlin spells out this conviction.

> The theory of time's passage I defend focuses on the *B* series: all events are ordered by a transitive, asymmetrical relation of earlier and later. Given events ordered in a *B* series, one can define an infinitude of *different A* series that correspond to taking different events as 'now' or 'present'. McTaggart's argument is marred throughout by his use of the phrase 'the *A* series' when there are, in fact, an infinitude of such. (MWP, note 11, p. 126)

This creates a tension, for it is hard to understand what passage amounts to, in a world where presentness, pastness and futurity are comparative, rather than absolute, properties. If time really is now flowing, some time from now will not be the same time as it (really) is now. Conversely, if, at all times, it really is the time it is at that time, but no time is singled out as "the present", then, prima facie at least, there is no sense in which time really flows. I will get back to this point. As Maudlin never explicitly mentions eternalism as such, before discussing how he intends to resolve this tension, let me first say why it is reasonable to assume that he is one.

> I know what it would be like to believe that the past is unreal (i.e. nothing ever happened, everything was just created *ex nihilo*) and to believe that the future is unreal (i.e. all will end, I will not exist tomorrow, I have no future). Insofar as belief in the reality of the past and the future constitutes a belief in a 'block universe', I believe in a block universe. (MWP, p. 109)

Let us set aside for a moment the uncharitable depiction of presentism as entailing that we have no future and that nothing ever happened (it takes for granted that the so called "truthmaker problem" cannot be solved). Maudlin's position seems to imply that the total state of the universe does not depend on what time it really is. Shortly afterwards, however, one reads that Maudlin is «one of those unusual defenders of the block universe who does not deny that there is an objective flow of time» (ibidem). The rest of this essay will be dedicated to an assessment of the author's attempted dissolution of this tension. While I think that Maudlin's views on time are deeply intriguing and thought provoking, I think that he has not done enough to dissolve it.

The analysis begins by noticing that the issue of passage is deeply connected to the issue of the directionality of time: the fact that there is an objective distinction between two directions in time, from any event, one towards the future (of that event), and one towards its past. Such directionality, as Maudlin rightly notices, cannot be accounted for by some intrinsic asymmetry of time. Space could well contain some intrinsic asymmetry, but this fact alone would not justify the claim that "space passes". Nor can the asymmetry in question be reduced to some contingent physical regularity associated with the directionality of time. It has often been suggested, for example, that the second law of thermodynamics, which implies that entropy never decreases, could be used to explain and reduce the directionality of time itself. According to these suggestions, the future direction of time is nothing but the direction in the fourdimensional manifold in which global entropy increases. The passage of time would amount to the fact that the entropy gradient is not zero. Maudlin argues (correctly, I think), that these reductive accounts cannot succeed. The second law of thermodynamics can only establish a (contingent) correlation between the entropy gradient and the directionality of time. The law was discovered and can be empirically tested. If the direction of time was not primitive (relative to the entropy gradient), how could we even express it? How would the world look like, if the second law did not hold? It would be a world, for example, where broken glasses would spontaneously recombine to end on the tables from where they fell. In order to even express this possibility, we have to assume that the "broken glass state" is instantiated *before* the "on the table state". If time itself were reversed by such an imagined reversal of the entropy gradient, then that world would be just indistinguishable from ours! In fact, the second law of thermodynamics would hold in that world, contrary to the hypothesis: in order to even express the

hypothesis that entropy never decreases, one has to first assume the directionality of time. As Maudlin quite aptly puts it, if the future-directed arrow of time amounted to nothing but the direction in which entropy increases, «[entropy] could go up and down like the stock market, but since the 'direction of time' would obligingly flip along with the entropy changes, entropy would still never decrease» (MWP, p. 129).

Lovers of desert landscapes have produced a great number of skeptical arguments purported to show that the passage of time is an illusion. In order to clear the ground for the thesis that passage is an essential ingredient in standard scientific explanations, Maudlin devotes a lot of effort (quite successfully, I think) to their debunkment. Here I will only touch on two of them. A standard argument against passage – see for example (Price, 1996) – stems from the observation that if it made sense to say that time passes it would also make sense to ask how fast it does so. But it clearly doesn't make sense to ask how fast time flows, for the only sensible answer would be that it flows at a rate of one second per second, which – so goes the argument – is nonsense. A rate of a second per second is nonsense because it is not even a rate: the units (seconds, in our example) cancel out to yield a dimensionless number, as it were. Maudlin contends, rightly, that units don't "cancel out" as suggested by the argument. $\prod$, for example, is defined as the ratio of a length to a length, but the units don't cancel out, to yield a dimensionless number. Sure there exists a real number, also called $\prod$, that stands in the same relation to unity as the circumference of a Euclidian circle to its diameter. But this does not impinge on the fact that the (geometric) definition of $\prod$ pertains to lengths, rather than temperatures or times. The real number called $\prod$ would exist even if the ratio of Euclidian circumferences to their diameters were a different number. Similarly, argues Maudlin, «the rate of passage of time at one second per second is still a rate: it, unlike $\prod$, is a measure of how much something changes per unit time» (MWP, p. 113–14).

The other standard skeptical argument that I wish to touch on derives from the alleged Time Reversal Invariance of the fundamental laws of physics. If the fundamental laws of physics are time reversal invariant, so goes the argument, then it follows that at the fundamental level our best understanding of nature does not distinguish the future direction from the past direction of time. Directionality would not be an intrinsic feature of spacetime, but a contingent fact about the distribution of matter. This picture is often the basis for those

accounts that we discussed above that attempt to reduce directionality to entropy gradients.

Maudlin's response to these arguments is twofold. On the one hand, he observes that the thesis that the laws of physics are time reversal invariant is false. According to the so called CPT theorem, observes the author, any quantum theory would be invariant under parity-plus-charge-conjugation-plus-time-reversal. It follows that the notorious violations of CP entail a violation of T. I find this response very intriguing, but unfortunately it is not explored any further in the book, so it is hard to assess its relevance to the issue at hand. Maudlin's second response is, I think, more compelling. Laws of nature alone, he observes, explain nearly nothing.

> The models of fundamental physical laws are infinitely varied, and the only facts that those laws alone could account for are facts shared in common by all the models. In all practical cases, we explain things physically not merely by invoking the laws, but also by invoking boundary conditions. (MWP, p. 120)

I think Maudlin is right on this point. The mere fact that the fundamental laws are time reversal invariant, even if it were true, would not suffice to prove that the distinction between the future direction and the past direction is not an intrinsic feature of time. To prove this one would also have to show that this distinction doesn't play any role in how laws are used, together with boundary conditions, to explain later states of the universe. I think Maudlin has powerfully argued to the contrary: if one assumes that nature is "blind" to the directionality of time, then all productive physical explanations lose their force. So if physics teaches us something about the passage of time, it is that passage is a real, intrinsic feature of spacetime.

Let me now turn to what I think is the most underdeveloped aspect of Maudlin's proposal: the dissolution of the tension between his brand of static eternalism and his realist views about the flow of time. While, as we said, the author puts a lot of effort in convincing us that "time passes", he is also very keen to reassure us that «except in a metaphorical sense, time does not move or flow» (MWP, p. 110). I don't quite feel reassured by Maudlin's stance on this point. Here is why. As we have seen, the feature of time that Maudlin argues is an essential ingredient in standard productive explanations, the feature that makes him claim that time really passes, is its directionality. Now, surely «the passage of time is *deeply connected* to the problem of the direction of time» (MWP, p. 109, my emphasis), but the directionality cannot be *identified* with

the passing of time itself. Indeed, Maudlin himself often claims that the directionality is "*produced* by passage", so he appears to take for granted that passage and directionality are not the same thing. What is passage, then, according to Maudlin? Because he takes passage to be a «fundamental, irreducible fact about the spatio-temporal structure of the world» (MWP, p. 107), Maudlin refuses to answer our question: «I cannot explain what I mean by paraphrasing or analyzing the notion of time's passage in terms that do not already presuppose the notion» (ibidem). Moreover, as we said, Maudlin is committed to a B-theory of time. In a literal sense, he thinks time does not flow at all. He dismisses the *locus classicus* of dynamical conceptions of time, the first part of McTaggart's famous argument, as a «mare's nest of confusions» (MWP, note 11, p. 126).

Nevertheless, I think that the use Maudlin makes of the undefined notion of "passage" exposes an unresolved tension. Like everyone else, Maudlin would like to have all the advantages of a dynamical conception of change without any of its notorious drawbacks. But the only relative "advantage" of his account is his refusal to conceptualize the nature of passage, hence of becoming. Let me explain with a few examples why I think that this strategy doesn't work. As I said, at the beginning of the chapter dedicated to the passing of time (chapter 4), Maudlin clears the ground from standard objections. Among them, we have seen, is the problematic status of the rate of passage. It is argued that these skeptical arguments miss their target, for it is perfectly fine to claim that time flows at a rate of one second per second. But if it is fine to claim that time flows at that rate, then surely it must be fine to assume that, at any given moment, time is found in a unique state as to which events are past, present and future. If it is possible to say that a train moves at a certain speed at a certain time, it must *eo ipso* be possible to say where the train is at that time. Likewise, if there is change as to the position of events with respect to their being past, present or future, then these events must have a precise position at any given moment to start with.

Indeed, in defending his view from the "no rate objection", Maudlin is forced to implicitly concede this point.

> If we ask how fast time flows, [...] we must mean to ask how the temporal state of things will have changed after a certain period of time has passed. In one hour's time, for example, how will my temporal position have changed? Clearly, I will be one hour further into the future, one hour closer to my death, one hour further from my birth. (MWP, p. 112)

Now, what can Maudlin possibly mean by "the temporal state of things", if not (something equivalent to) their position in McTaggart's A-series? If your "temporal state" is changing, so that in an hour's time it will be necessarily different from what it is now, then there must be a relevant temporal property that you possess right now, that can be acquired (you didn't have it an hour ago), and shed (you won't have it an hour from now). Moreover, such temporal property must be instantiated by the whole world, if it is to account for productive explanations that involve the evolution of the whole universe, as Maudlin would like.

"Fourdimentionalist", or "static" accounts of change have notoriously been accused of denying "genuine change". In particular, the notion of (comparative) change that static accounts can afford, essentially the possession of different properties at different times, has been argued to be unsuitable to feature in standard explanatory patterns. Russell's famous at-at theory of motion, for example, has been argued to be inadequate for the purposes that are usually assigned to change and motion in standard scientific explanations. These objections to the static conception of change go under the name of "no change objections". Now, Maudlin, on the one hand, as we have discussed, raises a brand of no change objection against those accounts of time that deny passage. Basically, he argues (correctly, I think) that without passage there is no directionality, and without directionality, no productive explanations. «The flow of the Missisipi [...] consist[s] in more than just the collections of instantaneous states that have different relative positions of the waters of the Missisipi to the banks» (MWP, p. 110). On the other hand, however, he would be happy to buy directionality, this "intrinsic asymmetry of the structure of time that has no spatial counterpart", without thereby committing himself to the idea that the total temporal state of the universe changes as times goes by. This, I argue, cannot be done. Note that Maudlin acknowledges that «the passage of time connotes more than just an intrinsic asymmetry: not any asymmetry would produce passing» (MWP, p. 129). So, the ingredient that is said to be missing from the Lewesian picture of time is some mysterious feature of the manifold of which we know nothing but the fact that it "produces" (in some unspecified way) the passing of time. Notice further that also the passing itself that is so mysteriously produced cannot be conceptually analyzed. All we know is that the mysterious feature of the manifold produces an undefinable aspect of time that we call "passage", and this, in turn, provides

the manifold with a direction that we call past-to-future, which we then use in our productive explanations.

Now, as Maudlin refuses to tell us what passage consists in, he can claim that, whatever it is, it does not entail the notion of an absolute present. Sure, if we introduce an unanalyzable property, we can decide, by *fiat*, all its conceptual relations to all other more familiar notions. So Maudlin is (formally) entitled to claim that "passage" is compatible with a world that lacks an absolute notion of presentness. But what entitles him to call it "passage"? Could we not just call it: "feature of the manifold produced by some peculiar asymmetry and responsible for the force of productive explanations"? No, we couldn't, because if we did we would turn this feature into a *virtus dormitiva*. What we find convincing and non circular about Maudlin's arguments, as far as the structure of time is concerned, is due to some pre-theoretic conceptual property of passage that we find intuitively responsible for the force of productive explanations. I suspect that this feature is the notion of a shifting present that casts its light on the various stages of the universe, one at the time, starting from the beginning to reach the present. Conversely, I think that Maudlin's block universe, once deprived of a shifting present, does not have the conceptual resources that are needed to run the arguments in favor of the non-Humean package.

REFERENCES

Hall, N. (2004). Two Concepts of Causation. In J. Collins, N. Hall, and L.A. Paul (Eds). *Causation and Counterfactuals*. Cambridge , MA: The MIT Press, 225–276.

Lewis, D. (1986). *Philosophical Papers, Vol. 2*. Oxford, UK: Oxford University Press.

McTaggart, J. M. E. (1927). *The Nature of Existence*. Cambridge, UK: Cambridge University Press.

Price, H. (1996). *Time's Arrow and Archimedes' Point*. Oxford, UK: Oxford University Press.

Suárez, M. (2009). The Many Metaphysics Within Physics. Essay Review of 'The Metaphysics Within Physics' by Tim Maudlin. *Studies in History and Philosophy of Science*, *Part B*, *40*(3), 273–276.

Commentary

Verity Harte
Plato on Parts and Wholes
Clarendon Press, Oxford 2002

Laura M. Castelli[†]
laura.castelli@exeter.ox.ac.uk

Verity Harte's book[1] proposes a reading of a series of interesting passages from Plato's Dialogues, where Plato sets forth different considerations about the way in which ontological unity and plurality can – or cannot – be conciliated in some form of unified ontological complexity. Harte reads the passages in a systematic way and this allows her to extrapolate the basic elements of a theory of parts and wholes, which she compares with two main contemporary positions (i.e. those of David Lewis and of Peter Van Inwagen[2]) on the problem of composition. Accordingly, the book can be read from two perspectives: as a reconstruction of an aspect of the philosophy of Plato and as a proposal of an alternative approach to the problem of composition and to the notions of part and whole. I shall first sum up what I take to be the main aspects of Harte's reconstruction of Plato's theory of parts and wholes and then comment on two main points, the first concerning the historical and interpretative contribution of the book and the second concerning the theoretical proposal there outlined.

The problem at issue can be formulated in the following way. There is something (say: Socrates), which we regard as *one* object. This object has parts (say: limbs), which are *many*. The one object is, in some sense, the same as its many parts; so, the same thing is one and many. But this is absurd because the same thing cannot be both one and many – at least not without qualification.

[†] Exeter College, University of Oxford, U.K.

[1] V. Harte, Plato on Parts and Wholes. The Metaphysics of Structure, Clarendon Press, Oxford 2002 (repr. 2006).

[2] See, in particular, D. Lewis, Parts of Classes, Blackwell, Oxford 1991; P. Van Inwagen, Material Beings, Cornell University Press, Ithaca, New York 1990; Id., "Composition as Identity", Philosophical Perspectives, 8 (Logic and Language), pp. 207–220.

The question is: Is any qualification available? It should be stressed here that the problem is not – or not merely – a problem of counting objects, but a matter of accounting for the ontology of non-atomic entities. If we indicate what we regard as one as "whole" and what we regard as many as "parts", the general problem is to understand what ontological claims are at stake if we want to make sense of

(C) whole = parts[3].

What is a whole? What are parts? And what relation does "=" indicate?

Harte explores Plato's approach to this problem by individuating two complementary groups of texts. The first group (discussed in Ch. 2) consists in a series of texts where Plato discusses (without endorsing) the conception of composition as identity, i.e. a conception according to which = in (C) just establishes that what appears on the left hand side of = is *just the same* as what appears on the right hand side of it *in the sense that* it is not ontologically "richer" than parts. According to the account at issue, parts always pluralize. Plato presents a series of puzzles resting on this assumption. Accordingly, if we want to retain the idea that the whole, which appears on the left hand side, is something one and that parts are many, either we must say that the presence of parts makes the whole a plurality (i.e. many) or we must say that the whole, if it has to be one, has no parts at all. In order to provide a more satisfactory account of the fact that a whole can be regarded as one despite the plurality of its parts, in the second group of texts (Chs. 3–4) Plato would then suggest to identify the (one) whole not with its (many) parts, but with a contentful structure. In order to convey an identity, (C) should be replaced by (C*):

(C*) whole = parts + structure.

(C*) should not be interpreted in the sense that structure is an additional part of the whole of the same kind as its material parts[4]. Rather, the idea would be that parts (or contents) and structure are the result of ontological analysis by way of abstraction on the ontology of the whole. Structure is nothing that the whole or the parts of the whole have and might lack; the whole *is* the structure of its contents.

[3] All formulations of (C) and the like are mine and are not present as such in the book.
[4] Cf. Arist., Met. VII 17, 1041b11–33; cf. also Top. VI 13, 150a1 ff.

Contrary to Lewis and with Van Inwagen, structure is essential to a whole; contrary to Van Inwagen, the structure is not just essential to the whole, but is also essential to its parts, which do not exist as "things" independently from the whole.

Some features of this account of composition should be stressed before moving to a couple of particular points of Harte's reconstruction. (i) This theory of structure rests at least in part on the neat separation of the ontological relations involved in composition from identity and otherness. In particular, the kind of ontological relations that are involved in composition are peculiar to it and cannot be reduced to identity: The relations of part to whole and of whole to part are sharply distinguished from the relations of identity and otherness. (ii) In addition, composition is ontologically creative and not innocent. Although we can analyse the ontology of a whole in terms of two components (structure and contents), we should not think of structure and contents as two distinct parts of the whole. The imposition of structure on unstructured contents is an abstraction which should not induce us to think that contents are there as separate and independently identifiable things. (iii) In Plato's account structure has a normative character such that either a whole is the result of the imposition of a good structure on the appropriate contents or there is no whole at all. The main idea seems to be that if some items (which might resemble the parts of a whole) are put together in a disordered and unbalanced way, the result of their composition cannot be regarded as a unified whole. (iv) Given the interdependence of structure and contents, it should be clear that, according to Plato, composition is not unrestricted (not any plurality of disparate objects counts as a whole). (v) In addition, the interdependence of structure and contents leads to the negation of extensionalism: It is not the case that the same materials build the same whole if they are structured in a different way[5].

Given this sketchy presentation of what I take to be the main points of Harte's reconstruction, I should like to say something on the proposed theory

[5] If contents are structure-laden in such a way that they can only be structured in one way, the problem does not arise in the first place. But this does not seem to be the case for at least some of the examples that Harte considers. Harte underlines that the structure is often characterized in mathematical terms, as a numerical proportion. Clearly, such a structure can be applied to very different contents. In developing this theory of structure, one might wish to be more precise on the proper level of generality of the abstract description of the structure.

as an interpretation of an aspect of Plato's metaphysics and, more generally, as a theoretical proposal.

From the historical and interpretative point of view, Harte certainly selects an interesting set of texts and provides a careful analysis of what she takes to be their "underlying ontology". One can disagree with her on the interpretation of particular points[6] as well as on her methodological assumptions on the connections between the texts[7], but I shall not comment on this, both because I do not think that this would be the appropriate place for such a discussion and because I think that it is useful to take Harte's methodological assumptions for the sake of the argument. She is very clear on the fact that her goal is not to provide a general account of Plato's metaphysics but to focus on Plato's approach to structure and I shall confine to her account of this particular aspect.

Harte says repeatedly that it is not completely clear whether Plato intends to apply his account of wholes as contentful structures to types and/or tokens, although, with the exception of the *Timaeus*, which portrays the construction of a particular token (the visible world), the examples seem mainly to refer to types. It is not very clear to me whether Harte thinks that the distinction between types and tokens reflects any important ontological distinction in Plato's ontology or not. Harte says more than once that she is not concerned with "what happens to Forms" (p. 6) or that she does not know how her analysis might apply to the case of Forms (p. 270). This is, I think, problematic for her reconstruction – and not (or not just) for the general reason that one might wonder whether it is possible to give an account of fundamental pieces of Plato's metaphysics without taking Forms into account. Rather, I think that what happens to Forms is quite crucial for Harte's interpretation.

In the first place, Harte relies on some passages (p. 140 ff.) in which Plato talks about the large-scale structure of the domains of sciences. Even if we do not want to say that each particular science such as music or mathematics has Forms as its objects, one might wonder how the large-scale structure is instantiated by the domain of what Plato takes to be the eminent form of philosophical knowledge, namely dialectic – and I assume as a relatively

[6] For example, Soph. 244b6–245e2, commented upon on p. 100 ff., is very difficult to understand und in some points it is possible that the Greek text is corrupted (e.g. at 244d11–12).

[7] For example, one might want to reject the idea that there is a "mereological undercurrent" in the Parmenides.

uncontroversial fact that, even if dialectic can be exercised with reference to any object, the primary objects of dialectic are Forms.

In the second place, Harte comments very clearly (p. 50 ff.) on Socrates' claim in the *Parmenides* that the most problematic case of joint ascription to the same subject of unity and plurality would be the case of Forms. If Harte's reconstruction does not apply to Forms, it is not very clear in what sense the metaphysics of structure that she proposes could be regarded as a very basic aspect of Plato's metaphysics. In particular, it seems to me that, if Harte reconstruction did not apply to Forms[8], there would be a problem to distinguish Plato's theory of structure with respect to Forms – which I take to be a basic part of his ontology – from Lewis' approach to composition as it is presented by Harte. Let me explain why. On p. 276 Harte presents Lewis' approach to parts in contrast with Plato's approach in the following terms: "Lewis has what one might call an 'atomistic' approach to composition. By this I do not mean that Lewis is committed to the existence of atoms. What I mean is that Lewis approaches composition from the bottom up. One starts with things, which are candidate parts, as the building blocks of composition. And one builds up to composites from these things by taking various sets of things, which are more or less related to each other in various ways. But the various ways in which the things in question are related (including their composing something) seem somehow secondary to the things themselves". Now, it seems to me that this kind of bottom-up approach to composition is discussed by Aristotle[9], at least on some occasions, in the form of an objection to Plato or to Platonists, who claim that each Form is one and that we can define Forms through their genus and specific differences. For instance, say that we want to define (the Form of) Man. The Platonists at stake would say that Man is Rational Biped Animal. Each of Rational, Biped and Animal is a Form and, as such, it is one. Now, why should the formula "rational biped animal" pick up one unified Form (i.e. the Form of Man) and not three distinct things (i.e. Rational, Biped and Animal)? It would seem that Harte's proposed approach to

[8] It should be clear that I am not saying that Harte's reconstruction does not apply to Forms; I am just saying that, despite the fact that she does not want to address the topic, assessing whether the reconstruction applies to Forms is quite important for her interpretation.

[9] Cf. e.g. Arist., Met. VII 15, 1040a14–23; and the generalized form of the problem (i.e. without reference to the theory of Forms) in Met. VII 12, 1037b10–14. See also the argument used in Met. VII 13, 1039a3–14 to show that a substance cannot be made of actual substances, each of which is one in actuality.

composition might provide an answer to Aristotle's objection *if* it can be applied to Forms. But unless we explain whether and how this approach applies to the Forms, the basic problem of the beginning of the *Parmenides*, i.e. how we talk of Forms in terms of one and many, remains intact.

Let me now move to the theoretical proposal that emerges from Harte's enquiry. As Harte herself suggests, the normative character (iii above) of the structure of wholes is "a feature of the account that one might wish to detach" (p. 274). This might be the case not just for reasons of "ideology" of the theory, but, as Harte seems to imply, for reasons of generality. In fact, what is presented as Plato's theory of composition is far from being a theory of composition in general, if we want to say that there might be composite objects that do not comply with Plato's requirements for wholes. As Harte suggests, one might ask what a theory like this would do with heaps. Harte outlines three possible answers: 1) there might be different kinds of composition; 2) there might be different degrees of wholeness; 3) heaps are just not wholes and the theory does not apply to them and does not say anything about their nature and existence. Harte comments on 1) by saying that Plato does not say anything about other kinds of composition and this fact might induce us to think that this would not be his reply to the issue of heaps (or of "bad" composites). It seems to me that both in Plato and in Aristotle *holon* ("whole") is used to designate not a generic composite, but a very precise composite, with a precise structure and a precise organization of its parts. The paradigmatic example of a whole is that of the living being, whose parts are organically unified and serve the purpose of the whole. If this is correct, then Plato's account of the structure of wholes *cannot* be a general account of composition. If this is the case, then we should ask the further question whether composition is wholeness (in the sense of "whole" of Plato). If it is, then we have to do with a very restricted notion of composition. But I am not sure at all whether we have enough evidence to say that composition is actually regarded as the same as wholeness. I personally believe that all we can say is that Plato (as Aristotle) has a very precise and restricted notion of wholeness[10] (not necessarily of composition). Clarity on this point might be relevant in order to better characterize the rejection of unrestricted composition ((iv) above).

[10] Nonetheless, I believe that both Plato and Aristotle accept the possibility of degrees of unity.

To conclude I would like to go back to (C) and to the way in which Plato analyses the views of those who take (C) as an identity claim. Harte correctly stresses (p. 83 ff.) that one feature of Plato's analysis of this point is that the relation of parts and whole is regarded as similar in nature to the relation of quantitative entities such as numbers and measures. To give an idea of this, we should think of (C) as analogous in structure to:

(C') $6 = 4+2 = 3+3 = \ldots = 1+1+1+1+1+1$

(C'') $1\text{m} = 10\text{dm} = 100\text{cm} = 1000\text{mm}$.

Even if Harte does not press this aspect too close, I think that this introduces an interesting element in the debate: = in (C) would indicate a sort of quantitative equivalence, saying that we have the same amount of stuff on both sides of the relation.

The problem with this, according to Plato, is precisely that the list of parts or "measures", which appears on the right hand side of =, simply states the presence of many items and does not say anything on how these should be arranged to make something one, just because arrangement is not relevant at all in this kind of equivalence. Other people, with different metaphysical intuitions, might be perfectly happy with the situation suggested by the association of (C) with (C') and (C''). It should be stressed that the supporter of (C*) can make sense of (C) in terms of quantitative equivalence or extensional coincidence (in the sense that the whole is exactly "where" its parts are and vice versa). One reason why Plato and, after Plato, Aristotle are not happy with (C) alone and require (C*) is that, although both claims state some relation of *sameness*[11] and, in this sense, express a *symmetrical* relation, the right hand side of (C*) has an explanatory power on the ontological structure of the left hand side that (C) lacks. In this sense, (C*) introduces a fundamental epistemological *asymmetry*: The right hand side provides an account of the ontological structure of what is on the left hand side and, in this sense, has explanatory priority. I am not sure what part explanatory tenets play in the contemporary debate on composition, but they certainly play an important role in Plato's and Aristotle's approach to the problem.

[11] I use "sameness" to indicate a looser relation than strict identity. This seems to be necessary in any case, given that there cannot be strict identity between one item and many items. Cf. Lewis on composition as "analogous" to identity.

Commentary

Kathrin Koslicki
The Structure of Objects
Oxford: Oxford University Press, 2008

Tuomas E. Tahko †
tuomas.tahko@helsinki.fi

Kathrin Koslicki's *The Structure of Objects* is a defence of a neo-Aristotelian approach towards composition, focusing especially on ordinary material objects. The central claim of the book is that a structure-based mereology for ordinary material objects can be given in terms of a single relation of parthood. Combined with an independent commitment to a realist yet moderate ontology of kinds, Koslicki's neo-Aristotelian mereology produces an account of composition which is opposed to the Composition-as-Identity (CAI) model familiar from the work of David Lewis; she argues instead that wholes are not to be identified with their parts, the two are numerically distinct.

It is impossible to do justice to Koslicki's rich book in this short commentary, which is why I will, in the spirit of this special issue, focus on her critique of the Lewisian account of composition and analyse the advantages that she claims her neo-Aristotelian approach to have over the CAI model.

The book consists of an overview of the 'standard' conception of the composition of material objects, the Lewisian four-dimensional account, and Kit Fine's alternative neo-Aristotelian model, as well as an analysis of Plato's and Aristotle's views on composition. From this basis Koslicki sets off to defend her own approach, a middle ground between the deflationary conception of structure present in Plato's and Fine's accounts on one hand and Aristotle's stronger, teleological approach on the other hand. In addition to the formal, mereological description of her account, Koslicki also provides a defence of the underlying ontology of kinds, motivated independently of mereological considerations, and explicates her conception of structure with case studies involving logic, chemistry, music, and linguistics.

† Department of Philosophy, History, Culture and Art Studies, University of Helsinki.

I will not discuss all of these aspects of Koslicki's book, interesting though they are. Instead I will compare Koslicki's approach to composition with the Lewisian approach. I am sympathetic to Koslicki's account and her reservations concerning Lewis's (as well as Ted Sider's) model, but I will also raise some concerns about her arguments and discuss some challenges that the positive thesis faces.

Let us begin with Koslicki's critical survey of the Lewisian approach. Koslicki's main targets are the principle of Unrestricted Mereological Composition (UMC) and the CAI thesis[1], which she discusses in Chapters 2 and 3 of *The Structure of Objects*. As Koslicki explains, Lewis thinks that Classical Extensional Mereology (CEM) is the only genuine kind of mereological composition (see Simons, 1987, pp. 37–41). It is CEM's commitment to UMC which Koslicki, quite rightly, considers to be the most crucial element in Lewis's account. Since Lewis's original argument for UMC is very dense, Koslicki follows Sider's (2001, p. 123) well-known version of Lewis's argument, which is commonly known as the 'vagueness argument'.

The vagueness argument claims that if UMC were false, there would have to be two adjacent cases in a continuous series such that in one composition occurs, but in the other it does not. Further, the argument claims that there is no such continuous series with a sharp cut-off concerning composition. The typical examples include baldness and heaps, but any sorites series will do: the point is that in all cases of composition, it either definitely occurs, or definitely does not occur. However, unless we accept unrestricted composition, we would need some criteria to judge where the sharp cut-off between non-composition and composition lies. Here is a passage from Sider himself which may help to explicate the argument:

> If not every class has a fusion then there must be a restriction on composition. Moreover, the only plausible restrictions on composition would be vague ones. But there can be no vague restrictions on composition, because that would mean that whether composition occurs is sometimes vague. Therefore, every class has a fusion. (Sider, 2001, p. 121)

The somewhat counter-intuitive upshot is that even my nose and the Eiffel tower compose an object, or indeed the fusion of the upper half of a trout and the lower half of a turkey, i.e. Lewis's 'trout-turkey'.

[1] See also Einar Bohn's discussion of CAI in his commentary of *Parts of Classes* in the present issue.

Koslicki, who is a friend of restricted composition, challenges the idea that (in all cases) composition either definitely occurs or definitely does not occur. Lewis' defence of this idea (1986, p. 212), as Koslicki (p. 34) points out, is unsatisfactory: it is based on the assumption that parthood (or overlap) is not vague. But since the original argument for unrestricted composition concerns the question of whether composition can ever be vague, and since the mereological notion of composition is *defined* in terms of parthood, it appears to be circular to assume that parthood cannot be vague at the outset.

Sider's attempt to circumvent this problem relies on the non-vagueness of logic, which Koslicki also grants (p. 36). What Koslicki does not grant is that this non-vagueness of logic contains everything that we can say about e.g. the existential quantifier: we may agree on the *meaning* of the quantifier, but disagree about its *range*. Koslicki thinks that the proponent and the critic of UMC can very well disagree about what and how many things exist, that is, what the existential quantifier can be legitimately said to range over. If she is right, the real disagreement is over what it means to be an object (or fusion), and hence the circularity objection to Lewis's original formulation stands its ground.

I believe that Koslicki is on the right lines: Sider's novel formulation of the Lewisian vagueness argument has at least one questionable premise. However, Sider (2003) has replied to Koslicki's concern, and Koslicki (p. 39) acknowledges that, at least insofar that vagueness is merely *linguistic*, the critic of UMC faces a challenge because vagueness requires *precisifications*: «Wherever there is vagueness (of the type relevant to the argument, anyway), there must be different non-vague candidate meanings "in the neighborhood of" the vague term» (Sider, 2003, p. 137). The classic move here is to adopt a «relatively precise background language» so that one can describe the different precisifications without the threat of *ontological* vagueness (instead of mere linguistic vagueness). But, the argument continues, no such background language is available in the case of quantifiers. There is more to be said about Sider's argument, but I shall instead raise a challenge for Koslicki's position (and for anyone else who wishes to deny UMC).

The upshot of Koslicki's discussion is that the debate about what it means to be an object (or fusion) remains open. Although this result blunts the vagueness argument somewhat, the burden of proof would seem to remain on the critic of UMC, since the proponent of UMC *does* have a simple answer to the question concerning what it means to be an object: any mereological fusion

constitutes one. The greatest challenge for the denier of UMC is to provide a *positive* account of what it means to be an object. Of course, Koslicki's ultimate goal is to do just that (and we will look into this shortly), but since her account is based on an ontology of kinds that needs to be motivated independently, her positive account does not provide a direct reply to the vagueness argument. I do however think that the vagueness argument can be refuted in the lines of Merricks (2005) and Tahko (2009), namely, by identifying a sharp cut-off in continuous series in terms of emerging causal powers.

Let us now briefly consider Koslicki's (pp. 40 ff.) analysis of the Lewisian CAI thesis, which states that composition is a kind of, or at least analogous to, numerical identity. Accordingly, fusions such as the trout-turkey are supposed to be unproblematic. Since the CAI thesis has been criticised extensively, Koslicki does not spend much time with it, she simply points out that the claim that a commitment to mereological sums does not carry with it any further ontological commitment 'over and above' the constituent objects of that sum is suspect. Koslicki (p. 42) asks us to consider a world which contains two mereological atoms, a and b, and hence according to UMC also a further object c, namely the sum of a and b. Now, we can agree that c is numerically distinct from a and b, so if one is ontologically committed to the sum of a and b, namely c, then one is committed to a further object c. According to Koslicki, this further commitment is objectionable, whereas a proponent of unrestricted composition claims that it is harmless since this commitment is supposedly 'nothing over and above' the commitment to the constituents of the mereological sum.

While I think that Koslicki's case against CAI is very plausible, I wish to make one point here. Even if the phrase 'nothing over and above' is ill-chosen, it is not clear to me that mereological fusions in Lewis's sense in fact do carry much ontological weight. The reason for this – and why Koslicki and other critics of UMC and CAI might think otherwise – is that in the Lewisian model the meaning of 'object' carries much less ontological weight than it does for someone like Koslicki. Therefore, perhaps a more charitable reading of Lewis's 'nothing over and above' is in the lines of Armstrong's (1997, p. 12) 'no addition to being', that is, sums should not be considered to add to the furniture of the world since they are merely concatenations of mereological atoms. Admittedly, introducing such additional metaphors may not be particularly helpful, but Armstrong's metaphor does at least serve to emphasize

the idea that 'object' could be understood as a mere umbrella term for the sum of its parts rather than an addition to the furniture of the world. Be that as it may, I wish to dedicate the remaining space to Koslicki's own, neo-Aristotelian conception of composition.

Chapter 7 of *The Structure of Objects* is where Koslicki does the bulk of the work towards her neo-Aristotelian account of composition. Perhaps the most important claim of this approach is that «material objects have *formal* parts in addition to their ordinary material parts» (p. 168). What are these formal parts? Koslicki describes them as a *recipe* that specifies «a range of selection requirements which must be satisfied by an object's material components» (p. 197). These requirements may include for instance the spatio-temporal proximity and the manner of arrangement of the object's material components. This is no doubt the most interesting and controversial part of Koslicki's account, so I will devote the rest of my discussion to it. These formal components of objects are also what determines when we have a genuine, successful case of composition – for Koslicki, what it means to be an object is that the recipe of a given object is satisfied by a selection of material components. Although the view is certainly controversial in postulating non-material parts, the concept of a recipe[2] is intuitively appealing: there are some criteria to judge when a set of material components composes an object of a certain kind, and the *arrangement* of those components according to a given recipe is crucial for an object of that particular kind.

Another point of interest in Koslicki's position is that she takes the formal components of objects to be proper parts of their respective wholes. The driving idea behind this is that any genuine kind of object has a set of formal proper parts, which act as the recipe according to which the relevant material components compose a whole of that particular kind. From these elements we get Koslicki's Neo-Aristotelian Thesis (p. 181):

> (NAT) Neo-Aristotelian Thesis: The material and formal components of a mereologically complex object are proper parts of the whole they compose.

[2] Koslicki's concept of a 'recipe' is metaphorical and encompasses three constraints that are associated with the *kind* that an object belongs to. These include (i) constraints concerning the *types* of material components of the object, (ii) constraints concerning the *arrangement* or *configuration* of the material components composing the object, and in some cases (such as water molecules), (iii) constraints concerning the *number* of material components of which a given whole must be composed.

The immediate advantage of NAT over the Lewisian line is that we can rule out gerrymandered fusions such as the trout-turkey: there are no recipes for such objects.

I will not discuss the details of Koslicki's argument for NAT here[3], rather, I wish to examine the general motivation for this view, which stems from an ontology of kinds. The commitment to an ontology of kinds is apparent in what Koslicki calls the Restricted Composition Principle (RCP, p. 173): a set of objects composes a further object of a particular kind just in case the original set of objects satisfies the formal constraints associated with that kind. Importantly, RCP is only appealing to those who are willing to accept that there are genuine natural kinds from which the formal constraints imposed on their composite objects emerge. Koslicki defends her own commitment to genuine kinds in Chapter 8 of *The Structure of Objects*, but I believe that there may be an interesting argument available to her even without a lengthy discussion of the ontology of kinds.

The argument that I have in mind (although I do not necessarily wish to commit to it myself) goes as follows:

(1) There is at least one genuine natural kind.
(2) Any genuine natural kind imposes formal constraints for its composite objects.
(3) Hence, at least one object has formal parts in addition to its material parts.

(3) follows from (1) and (2) given Koslicki's account of formal constraints (in Chapter 7 of *The Structure of Objects*). This argument, if correct, implies that NAT must be true of at least one object. From this result it is not difficult to extrapolate that NAT is probably true of many other objects as well, insofar as there is more than one genuine natural kind. Hence, only a thorough nihilist about natural kinds could deny NAT altogether (because she would deny the first premise). Since such nihilism is not commonplace, we have good reasons to take NAT quite seriously. Proponents of UMC will thus find themselves committed to not just one but two fairly counter-intuitive theses, namely the existence of things like trout-turkeys as well as the lack of genuine natural kinds.

Koslicki herself defends a much stronger account of natural kinds, but she does motivate the account with similar considerations, even though her

[3] But see (Bennett, 2011) for a reconstruction and a critique of this argument.

emphasis is on 'extra-mereological' considerations (pp. 233–234). In particular, she notes that the combination of an ontology of natural kinds and NAT produces a much more rigorous picture about the nature of wholes: in the Lewisian picture where everything, including trout-turkeys, goes, the existence of composite objects is motivated only by UMC and standard mereology. This has the unfortunate consequence of producing 'pseudo-kinds', whereas in Koslicki's neo-Aristotelian picture the existence of wholes of a particular kind can be motivated independently of mereological considerations. The upshot is a much more usable and intuitively plausible account of what it means to be an object.

It may be that Koslicki's positive picture is more appealing to those who are sympathetic to a neo-Aristotelian, realist ontology of kinds to begin with, but she does a good job in pointing out the advantages of the combination of this ontology of kinds with a neo-Aristotelian view of composition over the standard, Lewisian picture. The commitment to non-material formal parts that is central to this account can be regarded as an ontological cost, but I do believe it to be a necessary cost; I for one am more open to non-material formal parts rather than all manner of gerrymandered objects.

REFERENCES

Armstrong, D. M. (1997). *A World of States-of-Affairs.* Cambridge: Cambridge University Press.

Bennett, K. (2011), Koslicki on Formal Proper Parts. *Analysis, 71*(2), 286–90.

Lewis, D. (1986). *On the Plurality of Worlds.* Oxford: Blackwell.

Merricks, T. (2005). Composition and Vagueness. *Mind, 114*, 615–37.

Sider, T. (2001). *Four-Dimensionalism: An Ontology of Persistence and Time*. Oxford: Clarendon Press.

Sider, T. (2003). Against Vague Existence. *Philosophical Studies, 114*(1–2), 135–46.

Simons, P. (1987). *Parts: A Study in Ontology.* Oxford: Clarendon Press.

Tahko, T. E. (2009). Against the Vagueness Argument. *Philosophia, 37*(2), 335–40.

Commentary

Ted Sider
Temporal Parts

In T. Sider, J. Hawthorne, & D. Zimmerman (eds.), Contemporary Debates in Metaphysics. Oxford: Blackwell

Daniele Chiffi[†]
daniele.chiffi@unipd.it
Silvia Gaio[‡]
silvia.gaio@unipd.it

1.

The problem of change and the question of material constitution are two key ingredients of classical and contemporary debates in metaphysics and they are the main themes of this paper.

Sider has presented three arguments for a B-theory version of temporal parts theory. The idea behind temporal parts is that one object persists through time as well as it persists through space, that is, different properties can be applied to an object at different places, as well as different properties can be applied to an object at different times. According to McTaggart (1908), events in the B-series can be ordered according to the relational predicates "comes before" and "comes after", while the A-series implies a distinction between the non-relational predicates: "being past", "being present" and "being future". It is well-known that the B-series is a tenseless view about the nature of time, while the A-series is not. Usually the B-theory of time is associated to the ontology of temporal parts. Sider argues that the B-theory of time is equivalent to *eternalism* plus the thesis that "now" is like "here" inasmuch as they both have an indexical function. So, time is like space, and temporal parts exist as well as spatial parts. In case of future events, *eternalists* seem to hold the view that future already exists and deny the intuitive idea that future is an open

[†] Laboratory of Epidemiological Methods and Biostatistics, University of Padova, Italy.
[‡] University of Padova, Italy.

temporal dimension. The opposite view to eternalism is *presentism*, i.e. the idea that only present objects exist, while objects in the past and in the future do not exist.

Sider adds some examples in order to support the thesis that time is like space, notably regarding the reality of distant objects. Imagine that an object exists on Mars; although that object is spatially far from our actual world, nevertheless one can say that it exists. According to Sider, the same idea can be applied to temporally distant objects, i.e. even if the Jurassic era is temporally far from us and clearly the existence of dinosaurs cannot be physically tested, nevertheless dinosaurs exist to the same extent as the present objects that we perceive.

However, we point out that there are other relevant cases which are not handled by Sider's examples. Consider, for instance, a star in a distant region of the universe for which we do not have any physically direct proof. In this case, it is worth noting that the way to prove the existence of such a star is different from the ones for the dinosaur and the object on Mars. In our view it is paramount to draw some distinctions between different kinds of empirical proofs in order to refine the alleged similarity between time and space. A proof is *actual* if it involves a conclusive verification, while it is *potential* if it is in principle available. A proof is *direct* if what is proven can be ostensively shown, while it is *indirect* if it requires technical instruments of observation. In Sider's examples there is no direct actual proof involved, but the objects of his examples may allow an actual and indirect proof, while it is not the case for the example we provided (see Table 1).

In the next section we will introduce the three arguments for temporal parts provided by Sider.

2.

The first argument for temporal parts is called *argument from temporary intrinsics*. Such an argument is due to David Lewis (1986) who maintains that certain properties like shape are *intrinsic* since they do not require the existence of further objects in order to be fulfilled, while other properties require a relation with other objects. The latter can also be named *extrinsic* properties. As a way of example, "being round" for a table is an intrinsic property, while "being on the right of the chair" is an extrinsic property.

Lewis holds such a view because of the following metaphysical intuition: shape-like properties are intrinsic, for the instantiation of shape-like properties is not relative to time. If so, then shape-like properties have to be explained in terms of properties of temporal parts. Sider makes a good point observing that accepting extrinsic properties does not lead to a holistic perspective on the existence of objects in the external world since it does not require that *every* object is interconnected with all the other existing ones. Hence, the ontology based on extrinsic properties is not so odd.

How much reliable is Lewis's metaphysical intuition? In our opinion everyone who is acquainted with Abbott's novel *Flatland* can show good arguments against Lewis's intuition. Shape-like properties depend on your choice of geometry. In a flat land where nobody can observe the third dimension, all spheres will be considered to be like circles. This example shows that shape-like properties are not intrinsic since they may vary in different geometrical settings. If so, the distinction between intrinsic and extrinsic properties merely based on the existence (or not) of a relation with other objects is not reliable. Despite the unreliability of Lewis's intuition, one may hold temporal parts theory without the problematic distinction between intrinsic and extrinsic properties.

We switch now to the second argument for temporal parts which is based on the *paradox of material constitution*. Such a paradox arises when two different objects intuitively seem to be spatially coincident but yet related to their parts in different way. One example of the paradox of the material constitution is the following: imagine a sculptor that begins to shape a piece of clay on Monday, and she finishes her job on Tuesday. Now consider the paradox presented by Sider (2008):

> P1: the piece of clay that existed Monday continues to exist on Tuesday after being given statue shape.
> P2: The sculptor creates a statue, which exist on Tuesday but not on Monday.
> P3: If P1 and P2 are correct, then the statue and the piece of clay are two different material objects that on Tuesday are made up of exactly the same matter. They are not the same object because of Leibniz's Law: the piece of clay, but not the statue, exists on Monday.
> C: Therefore, different material objects can be made up of the same matter at a single time.

This argument comes to be paradoxical because the premises (P_1, P_2, P_3) are apparently sound while the conclusion C seems unacceptable. A further

example is the following: Consider now a certain cat, Tibbles, and its torso, Tib. Unfortunately for Tibbles, on Tuesday its tail is chopped off and the tail's matter is destroyed. We now argue as follows:

> P4: Tibbles exists on Tuesday, since a cat can survive the destruction of its tail.
> P5: Tib exists on Tuesday, since chopping off the tail did not affect Tib at all; it merely removed an external object that was once attached to Tib.
> P6: If P4 and P5 are correct, then on Tuesday, Tibbles and Tib are two different material objects made of the same matter. They are not the same object because of Leibniz's Law: Tibbles, but not Tib, had a tail as a part before Tuesday.
> C: Therefore, different material objects can be made up of the same matter at a single time.

Many attempts to solve the problem of constitution have been proposed. As noted by Sider, each of them accepts or rejects different points in the examples presented above (see Table 2). The *Constitution view* merely accepts cohabitation – i.e. two different objects occupying (parts of) the same space – and does not provide a reply to the paradox. *Mereological essentialism* is the perspective according to which parts are essential to the whole and a new object cannot be made of the same matter of other pre-existing objects. In other words, the only object that exists is the quantity of matter. Hence, a statue is not created as a new object as it is a mere mode of presentation of the same essence (matter) of the piece of clay. According to the *dominant sort theory* there is always a sort that prevails upon the others. As a result, cohabitation is rejected. On Monday the only existing object is the piece of clay, while on Tuesday the only existing object is the statue. Another view on the problem of material constitution follows from *nihilism*. According to this perspective, the ultimate constituents of matter are elementary particles and there is no sort at all[1].

Finally the last view on the problem of material constitution is provided by *temporal parts theory*. According to this theory, held by Sider, objects are aggregates of temporal parts, also named "spacetime worms". For instance, the statue is the collection of temporal parts which are a proper subset of the temporal parts of the piece of clay, since the piece of clay existed before the constitution of the statue and exists also after the disintegration of the statue.

[1] Table 3 explains under how many sorts we trace objects according to the four perspectives.

Hence, the statue and the piece of clay are two different objects, because they contain different temporal parts, even if they cohabit in some of them.

Sider maintains the view that we need to take into account the following concepts: the *tracing scenario* and the *tracing target*. The tracing scenario consists both of a series of times and of the objects existing at each of those times (tracing objects). In the case of the statue and the piece of clay, the tracing times are those when the piece of clay is shaped into the statue, while the tracing objects at each moment are the parts of the piece of clay. The tracing target of a scenario is a certain object x iff: a) the scenario «contains exactly the moments at which x exists» (p. 158), and b) «at each moment of S [the scenario], x is exactly composed of the tracing objects for that moment» (p. 158). In the aforementioned example, the scenario contains all the moments at which the statue exists and at each moment the statue is composed of the tracing objects (the parts of the piece of clay). So, the tracing target of this scenario is the statue.

Sider's fundamental claim is: every tracing scenario has a target. According to him, anyone who denies that all scenarios have targets must draw a line somewhere, between the scenarios that have targets and those that do not. Such a line, though, turns out to be very difficult to draw, for two reasons. First, the line must not be anthropocentric, and second, there is to take into account a problem of vagueness (this is the third argument for temporal parts). In Sider's opinion, anyone who claims that only some scenarios have targets believes that only ordinary objects exist and only those can be targets. If so, then strange objects, such as temporal parts, do not exist and they are not tracing targets. Sider points out that such a view is anthropocentric, but he does not discuss further the reasons grounding this view. Moreover, he underlines that, since ordinary concepts such as *statue*, *piece of clay*, *person* are vague, there could be tracing scenarios in which it is vague whether objects such as statues, pieces of clay and persons exist. In such cases, it is unclear whether the scenarios have targets. This is the core of the problem of vagueness associated with tracing scenarios. Temporal parts theory does not have to face issues concerning anthropocentrism and vagueness: all scenarios are considered to have targets and, therefore, ordinary objects are upon a par with strange objects: no line needs to be drawn.

Note that, while supporters of ordinary objects theories usually endorse a theory on what counts for an entity to be an object, Sider provides no one. This lack could be, after all, seen as a shortcoming of his theory. A possible way out

could come from the discussion of the problem of unrestricted composition. In footnote 12, in fact, Sider points out that his fundamental claim is based on the doctrine of unrestricted composition that we will take into account in the next section.

3.

In this final section we will raise some objections to Sider's assumption of unrestricted composition, also known as *universalism*. First and foremost, we want to point out that it is necessary to make a distinction in the ontology of the external world between structured objects and unstructured ones. We will call "structured objects" the objects that do not exist before the process of their construction; viz., the statue does not preexist before it is constructed by following a procedure. On the contrary, the piece of clay preexists to any manipulation, therefore we will call it "unstructured object".

We borrowed this distinction from the different roles that definition has in *Intuitionism* and in the *Platonistic* views of mathematics. In intuitionistic systems the definition creates the object that we want to define, while in Platonism the definition does not create an object because the object preexists to the definition itself (Quine, 1948). Similarly, the statue is the only structured object, while the piece of clay is not. There can be some moments in which the statue and the piece of clay cohabit but this happens just in a weak sense of cohabitation. Note that accepting the cohabitation of a structured object with an unstructured one is different from supporting the constitution view that is insensitive to the paradox of material constitution.

Especially in the case of structured objects, universalism is not a convincing point of view, as one can make objects following a procedure of constitution. Therefore not any aggregations of parts can contribute to constitute an object, but only the ones that are coherent with the possible combinations determined by the rule governing the procedure and the possibility to follow it.

Sider provides an outstanding approach to the ontology of time. Nonetheless, we believe that the issues above are still in need of further discussion.

	Actual proof	Potential Proof
Direct proof		Object on Mars
Indirect proof	Object on Mars /Dinosaur	Dinosaur/Object on Mars/ Distant Star (it may depend on the status of physical knowledge)

Table 1.

	P1	P2	P3	P4	P5	P6	C
Constitution view	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mereological Essentialism	Yes	No	Yes	No	Yes	Yes	No
Dominant sort	No	Yes	Yes	Yes	No	Yes	No
Nihilism	No	No	No	No	No	No	No

Table 2.

Under how many sorts do we trace objects?

Constitution view	All sorts
Mereological essentialism	One sort: quantity of matter
Dominant sort	One sort: dominant sort (it varies cases by cases)
Nihilism	None

Table 3.

REFERENCES

Lewis, D. (1986). *On the Plurality of Worlds*. Oxford: Blackwell.

McTaggart, J.E. (1908). The Unreality of Time. *Mind, 17*, 457–474.

Quine, W.V.O. (1948). On What There Is. *Review of Metaphysics, 2*, 21–38.

Book Review

Metametaphysics. New Essays on the Foundations of Ontology*

Oxford University Press, 2009

Giulia Felappi†
giulia.felappi@sns.it

Every discipline has its own instruments and studying them is part of the discipline itself. In order to draw true conclusions, for instance, an astronomer should know the extent to which the lenses of a telescope magnify. Likewise, a good metaphysician should know which are her working tools and how they can be used. Famously, by means of a study of metaphysical instruments, Carnap[1] reached an antimetaphysical stance. According to him, modern logic made it possible to analyse the main metaphysical instrument, i.e. language, with the result that metaphysical questions were dissolved: some questions – like *Does the nothing nothing?* – were discovered to be ill-formed and so incapable of being answered, others – like *Are there numbers?* – were analysable as questions not about metaphysical objects, but about language itself – along the lines of *Do we want to introduce number-words in our language?*

Following Carnap, many analytic philosophers developed a robust antimetaphysical attitude. Then something happened. Thanks to the works by Kripke, Lewis, Fine, van Inwagen, Armstrong and many others, metaphysics entered a new flourishing era. Metaphysical instruments were still analysed in order to obtain positive metaphysical results – consider, just to mention one example, Kripke's insight that proper names, qua instruments of rigid designation and direct reference, are appropriate tools for speaking about metaphysical necessity and possibility – but the antimetaphysical stance was

* Chalmers, D.J., Manley, D., Wasserman, R., (Eds.), (2009), *Metametaphysics: New Essays on the Foundations of Ontology*. Oxford University Press.

† Scuola Normale Superiore, Pisa, Italy.

[1] See (Carnap, 1937, Carnap, 1956 and Carnap, 1959).

set aside. Today it is certainly anachronistic to dismiss metaphysics as anachronistic[2].

The book *Metametaphysics* is a collection of seventeen essays where some of the most outstanding contemporary metaphysicians debate issues not within metaphysics but about its tools. (i) Some discussions concern specific instruments employed in metaphysics. (ii) Others are more general and what is assessed is the very possibility of metaphysics.

(i) In the book a number of rather different topics are investigated: *existence*, *naturalness*, *possible worlds*, *analysis*, *abstraction*, *definition*, *simplicity*, *common sense*, *fundamentality*, *primitiveness*. In order to give a sense of how the discussion proceeds, I will focus on the notion of *existence*. When we say that numbers and chairs exist are we employing a unique notion of existence or different notions for different kinds of objects? What is the relationship between *existence* as employed in ordinary language and the existential quantifier? Is there only one existential quantifier or a variety of them? These questions are classical within analytic philosophy, and some historical references are made to works in this area by Quine, Ryle, Meinong, Putnam and even Heiddeger. But some proposals, Fine's and Sider's, seem especially innovative. Fine rejects the Quinean thesis, so dominant in contemporary metaphysics, that sentences like *There are numbers* should be analysed as $\exists x$ *(x is a number)*. He advances arguments purporting to show that this kind of analysis yields wrong results and renders existence questions too easily answerable for them to be the central task of metaphysics. Fine therefore suggests we also reject the standard Quinean view according to which *existence* should be analysed via an existential quantifier. According to Fine, *There are numbers* should be analysed as $\forall x$ *(x is a number)* $\rightarrow$ *(x is real)*. Are numbers real? To answer difficult questions like this is the proper task of metaphysics, the study of the structure of reality. Sider, instead, accepts the Quinean claim that sentences like *There are numbers* should be analysed via the existential quantifier. Confronted with the possibility of a plurality of existential quantifiers, Sider maintains that, even if there were quantifier variance, there would still be a privileged quantifier in metaphysics, one which

[2] See (Williamson, 2007, p. 19).

is more natural, insofar as it is the only one that carves reality at the joints. Sider appeals to the notion of *naturalness*, a notion abundantly used in contemporary metaphysics. But while naturalness is usually applied to properties, in the wake of Lewis's division of properties into natural and not, Sider makes the original move of applying that notion to linguistic terms and their meanings.

The authors in the book focus, following Carnap, on linguistic instruments. But contemporary metaphysicians employ also other kinds of tools (*thought-experiments*, *intuition*, even *empirical experiments*) and it is arguably because they use those kinds of tools that metaphysics is now so flourishing. Some metametaphysical questions that naturally arise seem therefore to be worth investigating: Are thought-experiments reliable? May philosophy be experimental? Are we allowed or obliged to trust our intuitions? None of these questions are raised in the book and no reference is made to another recent work in metametaphysics, Tim Williamson's *The philosophy of philosophy*, or to the contemporary debates concerning experimental philosophy, where the focus is precisely on those topics.

Moreover, something seems to be missing also in the analysis of language. After Carnap, many notions have emerged that clearly bear on metametaphysics. For example, the notion of *inscrutability of reference* and the *principle of charity* seem to be central to any understanding of what a metaphysical debate is. When we debate with a metaphysician about the persistence, through time, of a rabbit, are we speaking about a rabbit or a rabbit-temporal-part? How should we interpret him and why so? In the book these topics are assumed without discussion or clarification.

(ii) Most discussions in the book consist in a general scrutiny of metaphysical tools with the aim in view of assessing the very possibility of metaphysics: given that metaphysics uses the tools it does, is scepticism about metaphysics anachronistic or is still a live option? Three are the kinds of scepticism advanced and criticised: Chalmers, Yablo and Thommason (criticised by Sider) advance the claim that at least some metaphysical sentences lack truth value; Hirsch (criticised by Bennett, Eklund, Hawthorne and Sider), suggests that some metaphysical debates are merely verbal disputes; Bennett maintains that we cannot in principle have evidence enough to establish the truth value of some metaphysical sentences. The sceptical proposals are analysed from the perspective of understanding what a

metaphysical debate is really about: are we discovering truths about reality or just deciding how to use our words? Are we making theoretical assertions or just suggesting proposals?

Not surprisingly, constant reference is made to Carnap. Price's and Soames's contributions are more historical in character and are devoted to understanding Quine's criticism of Carnap. They both conclude that the criticism does not really undermine Carnap's scepticism, which still is, therefore, a viable option.

What is surprising is that no substantive reference is made to the Carnap of *The logical syntax of language*, where he presented at length his metametaphysics. On discussing what is the appropriate language for philosophy, Carnap suggested that all languages are equally acceptable. A language L which introduces abstract entities is on a par with another, M, which does not. The debate about abstract entities turns out to be, therefore, a debate about what language to prefer. If, for instance, we want to express some laws of physics, we should choose L rather than M, not because L expresses correctly something about reality, but just because L is more useful than M, given our purpose. As Price points out in the book, Carnap is not Lewis[3], who argued that since modal realism is *useful* we have very good reasons to take it as *true*. Lewis employed a tool common in contemporary metaphysics, another of the tools not discussed in the book, the *inference to the best explanation*. According to Lewis, since realism about possible worlds permits better explanations in philosophy than any other hypothesis concerning possible worlds, we have good reasons to take it as *true*. Carnap took even the thesis that metaphysical disputes are disputes about which language to adopt not as *true* or *correct*, but just as a *useful* proposal. He dissolved the level of truth and correctness. For him there was no room for any inference to the best explanation and therefore it seems incorrect to ask Carnap for arguments in support of the *truth* of his proposal[4].

The sceptical positions presented in the book lack this crucial aspect of the Carnapian metametaphysics: they are advanced as theses, not as proposals. Therefore the reader is entitled to look for arguments in support of their truth or falsity. For admission of the very authors, the positions are only partially

[3] See (Lewis, 1986, pp. vii; 3-4).

[4] Cf. 'Lectures on Carnap'(Quine, 1990), Lecture III: "Carnap's thesis is to be regarded not as a metaphysical conclusion, but as a syntactic decision".

supported by weak arguments. This brings out the programmatic character of the book: having shown that scepticism in metaphysics is not anachronistic and that so many are the notions to be investigated, the book also shows that much in this area still needs to be done.

REFERENCES

Carnap, R. (1937). *The logical syntax of language*. London: Kegan Paul Trench, Trubner & Co.

Carnap, R. (1956). Empiricism, Semantic and ontology. In R. Carnap. *Meaning and Necessity*. Chicago: University Press, 205–213.

Carnap, R. (1959). The Elimination of Metaphysics Through Logical Analysis of Language. In A.J. Ayer (Ed.), *Logical Positivism*. New York, NY: Simon & Schuster, 60–81.

Lewis, D. (1986). *On the Plurality of Worlds*. Oxford: Blackwell.

Quine, W. v. O. (1990). Lectures on Carnap. In R.Creath (Ed.) *Dear Carnap, Dear Van: The Quine-Carnap Correspondence and Related Work*. Berkeley: University of California Press, 45–103.

Williamson, T. (2007). *The Philosophy of philosophy*. Oxford: Blackwell.

Report

"Another World is Possible" Conference on David K. Lewis University of Urbino, Italy, 16 – 18 June 2011

Adriano Angelucci[†]
adriano.angelucci@alice.it

The conference in celebration of David K. Lewis was held in Urbino, Italy, from June 16 to June 18 2011, and was organized by the University of Urbino, Faculty of Humanities and Philosophy, Department Of Communication Studies, Department of Basic Sciences and Foundations, together with APhEx, Analytical and Philosophical Explanations. The conference hosted five invited speakers, namely Andrea Bottani (University of Bergamo), Sònia Roca-Royes (University of Stirling), John Collins (Columbia University), John Divers (University of Leeds), Vincenzo Fano (University of Urbino), and a number of contributed papers. Aim of the conference was to explore further implications of Lewis' long-lasting impact on the fields of logic and metaphysics beyond the ones already established, while at the same time offering to young philosophers the opportunity to exchange views on related topics and the results produced by their own researches.

David Kellogg Lewis (1941–2001), as it is well known, was one of the most influential American philosophers of the 20th Century. His brilliant academic career was made at Harvard, where he worked under the supervision of W.V.O. Quine, UCLA, and Princeton. His long lasting love for Australia brought him to visit the country almost every year starting from the early seventies to the untimely end of his *actual* life. He made fundamental contributions to metaphysics and philosophy of mind, as well as significant contributions to areas as disparate as philosophy of language, philosophy of mathematics, philosophy of science, decision theory, epistemology, meta-ethics and aesthetics. As to the metaphysical side of his endeavours, two highly influential positions must be recalled, namely a Hume-inspired reductionism

[†] University of Verona, Italy.

about the nomological known as 'Humean supervenience', and a stance toward modality known as 'modal realism', according to which the best account of modality would be one which posits the existence of a plurality of concrete possible worlds. In philosophy of mind he defended a new version of materialism which motivated the view currently known as 'analytic functionalism', and offered original accounts of mental content and phenomenal knowledge. Besides an irritatingly high number of papers, he published four books: *Convention* (1969), *Counterfactuals* (1973), *On the Plurality of Worlds* (1986), and *Parts of Classes* (1991). The occasion for the conference was provided by the 10th anniversary of David Lewis premature departure.

An attempt at providing an exhaustive account of all the talks delivered at the conference is way beyond the reach of this short review, which is rather intended to sketch a brief outline of the main theses defended by some of its contributors, and which will focus primarily on the contributions of young scholars.

Sònia Roca-Royes presented a paper titled *Conflation, Primitive Modality, and the Humean Intuition*. She explored some of the pros and cons of two different views concerning the nature of possible worlds, namely Linguistic Abstractism and Lewisian Concretism. While the former is the view according to which possible worlds are sets of maximally consistent propositions, and hence abstract entities, the latter holds that they are concrete entities, ontologically on a par with our actual world. In particular, she argued for the existence of an unavoidable internal tension in Lewisian Concretism, which would not be present in the *ersatz* accounts of modality. According to her view, Lewisian Concretism would be problematically committed to the presence of necessary connections between different existents. The existence of such necessary connections would in fact undermine that very same Human principle which Lewis relies on in order to support his Principle of Plenitude. This last principle, as a matter of fact, is crucial to his account, insofar as it confers a higher descriptive power to Concretism over Ersatzism. If she were to be right, then this would significantly weaken Lewis' general abductive argument in favour of Concretism.

John Collins's paper, titled *The Parsing of the Possible*, was intended to address the question of whether it is possible to carve the logical space of possibility more finely than the tools of counterfactual analysis permit. He started by reviewing fundamental arguments aimed at rejecting the conditional

analysis of dispositions, such as the arguments from *finkish* and *masking*. *Finkish* cases are, roughly, those in which the conditions for an object's acquiring or losing a disposition coincide with the conditions of its manifestation. *Masking* cases would be those in which the manifestation of the disposition is not removed but simply masked by something else. He then argued, contra Lewis, that rejecting *masked* dispositions would represent a significant loss of descriptive power for a satisfactory metaphysical account of component dispositions. Building on such premises he finally hinted toward a new counterfactual theory of causation.

Michael De (University of St. Andrews) presented his paper *A Modal Realist Defense of Presentism*, whose main aim was to show that modal realists have the resources required in order to avoid the major objections traditionally raised against Presentism. The objections which De decided to focus on are: the singular proposition objection, according to which there are singular propositions about past objects; the cross-temporal relation objection, according to which present objects can stand in specifiable relations to past objects; the causation objection, according to which the causes of events are past events; and the truth-making objection, according to which the truth-makers of some propositions are past objects or facts. In his view modal realists are in the position to consistently claim that the entities to which all these objections refer are not strictly past entities, but rather other-worldly ones.

Marco Nathan (Columbia University) presented a paper titled *Lewisian Themes in Molecular Biology*. His general aim was to show the relevance of Lewis' thought for the philosophy of science. Lewis' work on redundant causation and on the conditional analysis of dispositions, in his view, finds applications in current scientific practice. In order to show this, he discussed two examples from molecular biology, namely the operon model of gene regulation, and the instantiation of *finkish* dispositions and *masking* in some biological processes. Both examples, according to his view, respectively instantiate two central concepts of Lewisian metaphysics, namely *preemption* and *dispositions*. The moral to be drawn from these examples would be that philosophy of science and metaphysics are more closely related than is often assumed.

The above examples will be sufficient, we hope, to convey at least a rough idea of the deeply fascinating Lewisian themes dealt with during the conference both by young researchers and well-established authorities in the field. The highly original and far-reaching ideas of the great American

philosopher, it seems, have not ceased to inspire generations of new philosophers yet!

HUMANA.MENTE - ISSUE 19, DECEMBER 2011

Composition, Counterfactuals and Causation

Edited by M. Carrara, R. Ciuni, G. Lando

The problems of how the world is made, how things could have gone, and how causal relations work (if any such relation is at play) cross the entire historical development of philosophy.

In the last forty years, the analytic tradition has given these problems a prominent role in its agenda, and David Lewis has suggested methodologies and theories that have contributed to enrich the complexity of our main notions in the field of, mereology, modality and the theory of causation.

For more information about the journal
visit our website at:
www.humanamente.eu

www.humanamente.eu

Humana.Mente – Journal of Philosophical Studies was founded in Florence in 2007. It is a peer-reviewed international journal that publishes 4 issues a year. Each issue focuses on a specific theme, selected from among critical topics in the contemporary philosophical debate, and is edited by a specialist on the subject, usually an emerging researcher with both philosophical and scientific competence.

Humana.Mente wants to be a place for exploring the most recent trends in the international philosophical discussion and wants to give the opportunity to the international community of young researchers to confront each other, and to discuss, control and verify their theories. An analytic perspective is favored, and particular attention is given to the relationship between philosophy and science, without however neglecting the historical aspects of the philosophical topics.

In this issue works by

Andrea Iacona, Ned Hall, Henry Laycock, Massimiliano Carrara, Enrico Martino, Andrea Borghini, Giorgio Lando, Giuliano Torrengo, Einar Duenger Bøhn, Louis deRosset, Vittorio Morato, George Darby, Tim Maudlin, Emiliano Boccardi, Laura Castelli, Tuomas Tahko, Daniele Chiffi, Silvia Gaio

Humana.Mente
Journal of Philosophical Studies

Issn: 1972 -1293